Asymptotic distribution of estimators in reduced rank regression settings when the regressors are integrated

Dietmar Bauer *
Austrian Institute of Technology, Dynamic Transportation Systems
Giefingg. 2, A-1210 Wien

November 8, 2012

Abstract

In this paper the asymptotic distribution of estimators is derived in a general regression setting where rank restrictions on a submatrix of the coefficient matrix are imposed and the regressors can include stationary or I(1) processes. Such a setting occurs e.g. in factor models. Rates of convergence are derived and the asymptotic distribution is given for least squares estimators as well as fully-modified estimators. The gains in imposing the rank restrictions are investigated. A number of special cases are discussed including the Johansen results in the case of cointegrated VAR(p) processes.

Keywords: rank restricted regression, asymptotic distribution, integration

1 Introduction

In this paper a multivariable time series $(y_t)_{t\in\mathbb{Z}}, y_t \in \mathbb{R}^s$, is modeled as a linear function of two processes $(z_t^r)_{t\in\mathbb{Z}}, z_t^r \in \mathbb{R}^{m_r}$ and $(z_t^u)_{t\in\mathbb{Z}}, z_t^u \in \mathbb{R}^{m_u}$ (where 'r' stands for restricted and 'u' for unrestricted) using the following model:

$$y_t = b_r z_t^r + b_u z_t^u + u_t, t = 1, \dots, T \tag{1}$$

where $b_r = O\Gamma'$ is of rank $n < \min(s, m_r)$. Such a situation can occur e.g. for panel data sets where both s and m_r are large. Throughout all variables will be assumed to be either stationary or (co-)integrated. Details on the assumptions for the processes are given below. For the moment assume that $(u_t)_{t\in\mathbb{Z}}$ is an independent identically distributed (iid) process.

In this situation the asymptotics for the OLS estimators (Park and Phillips, 1988; Park and Phillips, 1989) and fully modified (Phillips, 1995) estimators neglecting the rank restriction are well documented in the literature. However, neglecting the rank restriction in the case

*Correspondence to: tel.: ++43 +50 550 6669, e-mail: Dietmar.Bauer@ait.ac.at.

that m_r and s are large, the number of parameters to be estimated equals $(m_r + m_u)s$ which might require excessively large samples in order to allow for reasonable accuracy. As an alternative then rank restricted regression (RRR) can be used in order to reduce the number of parameters greatly.

The RRR framework of equation (1) is also of importance for the estimation involved in subspace methods, (see e.g. Larimore, 1983; Bauer and Wagner, 2002). In these methods a RRR of the type (1) is the central step in the estimation. Thus the understanding of the asymptotic properties of the corresponding estimators needs a thorough understanding of the asymptotic properties of estimators for (1).

If all involved processes are stationary the asymptotic theory of RRR estimators based on OLS is presented in (Reinsel and Velu, 1998). There consistency and asymptotic normality of the estimated coefficient matrices is stated for the (generic) special case that all singular values of b_r are distinct. Further expressions for the asymptotic variance matrix are provided using implicitly defined quantities which, hence, are not easy to interpret or implement.

For a cointegrated process X_t letting

$$y_t = \Delta X_t = X_t - X_{t-1}, \quad z_t^r = X_{t-1}, \quad z_t^u = [\Delta X_{t-1}', \ldots, \Delta X_{t-p}']'$$

equation (1) corresponds to the Johansen framework (Johansen, 1995). Also in this case the asymptotics of quasi maximum likelihood estimators are well known. Although the original material focuses on the estimation of the cointegrating relations Γ extracted as the right factor in the product $b_r = O\Gamma'$, the asymptotics for the full estimator $\hat{b}_r$ can be derived based on these results, see the evaluations in (Johansen, 1995). The arguments given there rely on stationarity of y_t and $\Gamma' X_{t-1}$ as well as on the fact that the rank restriction only restricts the coefficients corresponding to the nonstationary components of X_{t-1} as will be demonstrated below.

Equation (1) extends this framework by allowing for more general processes z_t^r and z_t^u. It will be shown below (see Theorem 3.1) that there exist nonsingular transformations $\mathcal{T}_y, \mathcal{T}_r$ such that

$$\tilde{b}_r := \mathcal{T}_y b_r \mathcal{T}_r^{-1} = \begin{bmatrix} I_{c_y} & 0_{c_y \times (c_r - c_y)} & 0_{c_y \times (m_r - c_r)} \\ 0_{(s-c_y) \times c_y} & 0_{(s-c_y)\times(c_r - c_y)} & \tilde{b}_{23,r} \end{bmatrix} = \underbrace{\begin{bmatrix} I_{c_y} & 0 \\ 0 & \tilde{O}_2 \end{bmatrix}}_{\tilde{O}} \underbrace{\begin{bmatrix} I_{c_y} & 0 & 0 \\ 0 & 0 & \tilde{\Gamma}_{32}' \end{bmatrix}}_{\tilde{\Gamma}'}$$

and in $\tilde{z}_t^r = \mathcal{T}_r z_t^r$ the first c_r coordinates are integrated, the remaining ones being stationary. In the Johansen framework $c_y = 0$ holds while in this paper $0 \le c_y \le n \le s$ is allowed

for. Also in $\mathcal{T}_y(y_t - b_u z_t^u)$ the first c_y components are integrated the remaining ones being stationary.

In this extended situation the asymptotics of (Johansen, 1995) do not apply as can be seen from the following arguments: Using the notation $\langle a_t, b_t\rangle = T^{-1}\sum_{t=1}^T a_t b_t'$ for processes $(a_t)_{t\in\mathbb{Z}}, (b_t)_{t\in\mathbb{Z}}$ the consistency proof in Lemma 13.1. of (Johansen, 1995) relies on solving the generalized eigenvalue problem

$$\lambda\langle \tilde{z}_t^{r,\pi}, \tilde{z}_t^{r,\pi}\rangle v - \langle \tilde{z}_t^{r,\pi}, y_t^\pi\rangle\langle y_t^\pi, y_t^\pi\rangle^{-1}\langle y_t^\pi, \tilde{z}_t^{r,\pi}\rangle v = 0.$$

Here a_t^π denotes the residuals of a regression onto z_t^u. Consistency is shown by transforming the problem using the matrix $A_T = [\tilde{\Gamma}_\perp T^{-1/2}, \tilde{\Gamma}]$ (changing the order of the block column to correspond to our ordering as used below) where the columns of the matrix $\tilde{\Gamma}_\perp, (\tilde{\Gamma}_\perp)'\tilde{\Gamma}_\perp = I$ span the orthogonal complement of the space spanned by the columns of $\tilde{\Gamma}$. Correspondingly for $c_y = 0$ in $A_T' z_t^{r,\pi}$ the first components are nonstationary but scaled by $T^{-1/2}$ and the remaining ones stationary. In the transformed problem

$$\lambda A_T'\langle \tilde{z}_t^{r,\pi}, \tilde{z}_t^{r,\pi}\rangle A_T w - A_T'\langle \tilde{z}_t^{r,\pi}, y_t^\pi\rangle\langle y_t^\pi, y_t^\pi\rangle^{-1}\langle y_t^\pi, \tilde{z}_t^{r,\pi}\rangle A_T w = 0$$

all matrices converge to block diagonal matrices. Thus the corresponding eigenvalues and matrix of eigenvectors V_T (with a suitable choice of the basis) converge. The eigenvectors of the transformed problem corresponding to the nonzero eigenvalues are related via $V_T = A_T^{-1}W_T = [\tilde{\Gamma}, \tilde{\Gamma}_\perp T^{1/2}]' W_T$ (assuming without restriction of generality $\tilde{\Gamma}'\tilde{\Gamma} = I_n$) implying that $T^{1/2}\tilde{\Gamma}_\perp' W_T$ converges to zero in probability. No almost sure (a.s.) results and no sharper bounds on the order of convergence are provided in (Johansen, 1995).

For $c_y > 0$, however, $\tilde{\Gamma}'\tilde{z}_t^r$ is nonstationary and $A_T'\langle \tilde{z}_t^{r,\pi}, \tilde{z}_t^{r,\pi}\rangle A_T$ does not converge. Using instead $\tilde{A}_T$ as

$$\tilde{A}_T' = \begin{bmatrix} T^{-1/2}I_{c_y} & 0 & 0 \\ 0 & T^{-1/2}I_{c_r - c_y} & 0 \\ 0 & 0 & \tilde{\Gamma}_{32}' \\ 0 & 0 & \tilde{\Gamma}_{32,\perp}' \end{bmatrix}$$

where $\tilde{\Gamma}_{32,\perp}'\tilde{\Gamma}_{32} = 0, \tilde{\Gamma}_{32,\perp}'\tilde{\Gamma}_{32,\perp} = I$ leads to convergence for the generalized eigenvalue problem. Thus again $V_T = \tilde{A}_T^{-1}W_T$ converges, where the first c_y columns corresponding to the eigenvalue $\lambda = 1$ converge to $[I_{c_y}, 0]'$. Consequently also $W_T = \tilde{A}_T V_T$ converges. However, the heading $c_y \times c_y$ subblock of this matrix equals $T^{-1/2}I_{c_y}$ and hence converges to zero as does the whole block column. Multiplying the corresponding block column with $T^{1/2}$ the

heading subblock equals the identity matrix as required, but the orders of convergence for the remaining blocks are reduced by this order and hence the remaining arguments in the proof of Lemma 13.1 of (Johansen, 1995) do no longer apply. Therefore this approach cannot be used in order to show consistency for the estimator of $\tilde{\Gamma}$ and thus also not of $\tilde{O}$. Due to this complication (Bauer and Wagner, 2002) were led to provide an adapted estimator by setting the remaining block rows of the first block column of V_T equal to zero. In this paper a different route in the proof for consistency of the estimator for b_r is provided showing that the adaptation is not needed.

In addition to the changes in the consistency proofs also the derivation of the asymptotic distribution of the estimators $\check{O}$ and $\check{\Gamma}$ of $\tilde{O}$ and $\tilde{\Gamma}$ as provided in Lemma 13.2. of (Johansen, 1995) for the case $c_y = 0$ cannot be used in the case $c_y > 0$ as can be seen from these arguments: In the last equation on p. 182 the last term $(\check{O} - \tilde{O})\tilde{\Gamma}'\langle \tilde{z}_t^{r,\pi}, \tilde{z}_t^{r,\pi}\rangle$ is shown to tend to zero using consistency for $\check{O}$ and stationarity of $\tilde{\Gamma}'\tilde{z}_t^{r,\pi}$ for $c_y = 0$. For $c_y > 0$, however, $\tilde{\Gamma}'\tilde{z}_t^{r,\pi}$ contains nonstationary components such that $\langle \tilde{\Gamma}'\tilde{z}_t^{r,\pi}, \tilde{z}_t^{r,\pi}\rangle = O_P(T)$ and moreover converges in distribution to a nondegenerate distribution when divided by T. Hence even if $(\check{O} - \tilde{O})$ is estimated superconsistently such that $T(\check{O} - \tilde{O})$ converges in distribution, the term $(\check{O} - \tilde{O})\tilde{\Gamma}'\langle \tilde{z}_t^{r,\pi}, \tilde{z}_t^{r,\pi}\rangle$ in the last equation on p. 182 does not vanish. Thus also for the asymptotic distribution the proof in (Johansen, 1995) does not apply for the case $c_y > 0$ and a more detailed analysis is needed. It is the main goal of the paper to close this gap in the literature.

In this paper two different estimators are considered: RRR estimator based on the unrestricted OLS estimator as well as based on the fully modified unrestricted estimator of (Phillips, 1995). The main contributions of the paper are:

- A full discussion of the asymptotic properties of the RRR estimators including conditions for consistency, derivation of the asymptotic distribution of the estimators under the condition of known rank n is provided.
- For the RRR estimator based on OLS almost sure (a.s.) rates of convergence are provided, improving the results in the literature which provide only in probability convergence.
- Furthermore in all cases the asymptotic distribution will be given explicitly and a detailed comparison of the relative advantages in a number of special cases is provided.

The organization of this paper is the following: The next section presents the various estimation algorithms while their corresponding asymptotic properties are discussed in section 3. Section 4 illustrates the results using a number of special cases. Finally section 5 summarizes the paper. All results are proved in Appendix A. A summary of the notation is contained in Appendix B.

2 Estimation Algorithms

In this paper four different estimators for the coefficient matrices b_r, b_u in equation (1) based on observations for time instants $t = 1, \dots, T$ are considered. Throughout as above the notation $\langle a_t, b_t \rangle := T^{-1} \sum_{t=1}^T a_t b_t'$ will be used (somewhat sloppily using a_t, b_t for the processes $(a_t)_{t\in\mathbb{Z}}$ and $(b_t)_{t\in\mathbb{Z}}$ and for the variables a_t, b_t for given time instant t respectively).
Using this notation the ordinary least squares (OLS) estimator (that ignores the knowledge on the rank constraint $\text{rank}(b_r) = n$) can be written as

$$\hat{\beta}_{OLS} = \langle y_t, z_t \rangle \langle z_t, z_t \rangle^{-1}, \quad \hat{\beta}_{OLS} = [\hat{\beta}_{OLS,r}, \hat{\beta}_{OLS,u}].$$

If

$$\tilde{\Xi_+}(\tilde{\Xi_+})' := \langle y_t^\pi, y_t^\pi \rangle^{-1}, \quad y_t^\pi = y_t - \langle y_t, z_t^u \rangle \langle z_t^u, z_t^u \rangle^{-1} z_t^u,$$

the rank restricted estimator maximizing the quasi maximum likelihood based on the assumption of iid Gaussian residuals can be defined as

$$\hat{\beta}_{RRR} = \arg \min_{\beta=[\beta_r,\beta_u]\in\mathbb{R}^{s\times(m_r+m_u)},\text{rank}(\beta_r)=n} \text{tr} \left[\tilde{\Xi_+} \sum_{t=1}^T (y_t - \beta_r z_t^r - \beta_u z_t^u)(y_t - \beta_r z_t^r - \beta_u z_t^u)' \tilde{\Xi_+}' \right]$$

and is given by

$$\hat{O} = (\tilde{\Xi_+})^{-1} \hat{U}_n \hat{S}_n, \hat{G}' = \hat{V}_n' (\tilde{\Xi_p^-})^{-1}, \hat{\beta}_{RRR,r} = \hat{O}\hat{G}', \hat{\beta}_{RRR,u} = \langle y_t - \hat{\beta}_{RRR,r} z_t^r, z_t^u \rangle \langle z_t^u, z_t^u \rangle^{-1} \quad (2)$$

using the SVD

$$\tilde{\Xi_+} \hat{\beta}_{OLS,r} \tilde{\Xi_p^-} = \hat{U}_n \hat{S}_n \hat{V}_n' + \hat{R}_n$$

where $\hat{U}_n$ denotes the matrix having as columns the singular vectors corresponding to the dominant singular values $\hat{\sigma}_1 \geq \hat{\sigma}_2 \geq \ldots \geq \hat{\sigma}_n > 0$ contained as the diagonal in the diagonal matrix $\hat{S}_n$. The corresponding right singular vectors are contained in $\hat{V}_n$. Finally $\hat{R}_n$ constitutes the approximation error. Here $\tilde{\Xi_p^-} = \langle z_t^\pi, z_t^\pi \rangle^{1/2}$ (where $X^{1/2}$ denotes the

symmetric matrix square root of the square matrix X and z_t^π denote residuals from regression of z_t^r onto z_t^u). Clearly the estimator $\hat\beta_{RRR,r}$ does not depend on the decomposition of $\hat O\hat G'$ into $\hat O$ and $\hat G'$.

Note that for this choice of $\tilde{\Xi_+}$ and $\tilde{\Xi_p^-}$ the columns of $\hat G'$ can also be interpreted as the eigenvectors to the generalized eigenvalue problem

$$\langle z_t^\pi, z_t^\pi\rangle \hat G \hat S_n^2 = \langle z_t^\pi, y_t^\pi\rangle\langle y_t^\pi, y_t^\pi\rangle^{-1}\langle y_t^\pi, z_t^\pi\rangle \hat G.$$

As can be verified straightforwardly the corresponding estimate $\hat O$ equals the coefficients for regressing y_t^π onto $\hat G' z_t^\pi$. Thus in the Johansen framework the Johansen estimators are obtained.

(Phillips, 1995) discusses the fully-modified (FM) estimators as an alternative to least squares estimation. The fully modified OLS estimator (FM-OLS) of β is defined as

$$\hat\beta_{OLS}^+ := \Big(\langle y_t, z_t\rangle - \hat\Delta_{\hat u,\Delta z} - \hat\Omega_{\tilde u,\Delta z}\hat\Omega_{\Delta z,\Delta z}^{-1}(\langle \Delta z_t, z_t\rangle - \hat\Delta_{\Delta z,\Delta z})\Big)\langle z_t, z_t\rangle^{-1} \tag{3}$$

where for processes $(a_t)_{t\in\mathbb{Z}}$ and $(b_t)_{t\in\mathbb{Z}}$ the estimates

$$\hat\Omega_{a,b} := \sum_{j=1-T}^{T-1} w(j/K)\hat\Gamma_{a,b}(j), \quad \hat\Delta_{a,b} := \sum_{j=0}^{T-1} w(j/K)\hat\Gamma_{a,b}(j)$$

are used. As usual $\Delta z_t := z_t - z_{t-1}$. Here $\hat\Gamma_{a,b}(j) := \langle a_t, b_{t-j}\rangle = T^{-1}\sum_{t=1}^T a_t b_{t-j}'$ denotes the estimated covariance sequence where observations outside of the observed sample are treated as zeros. Further $\hat\Omega_{\tilde u,\Delta z}$ is estimated using the residuals $\hat u_t = y_t - \hat\beta_{OLS} z_t$. Throughout we will use the subscripts to indicate the processes involved. Additionally superscripts indicate components of the processes. A slight difference to the notation of e.g. Phillips (1995) is that for the integrated processes z_t, say, we index by Δz rather than only z.

Consequently for stationary processes $(a_t)_{t\in\mathbb{Z}}$ and $(b_t)_{t\in\mathbb{Z}}$ it follows that $\hat\Omega_{a,b}$ and $\hat\Delta_{a,b}$ are estimators of the long-run-covariance and the one-sided long run covariance matrices defined as

$$\Omega_{a,b} = \sum_{j=-\infty}^{\infty} \mathbb{E} a_j b_0', \quad \Delta_{a,b} = \sum_{j=0}^{\infty} \mathbb{E} a_j b_0'.$$

For the kernel function $w(\cdot)$ occurring in this definition we will use the standard assumptions (cf. Phillips, 1995):

Assumption K: The kernel function $w(.) : \mathbb{R} \to [-1, 1]$ is a twice continuously differentiable even function with

(a) $w(0) = 1, w'(0) = 0, w''(0) \neq 0$

(b) $w(x) = 0, |x| \geq 1$ with $\lim_{|x|\to 1} w(x)/(1-|x|)^2 =$ constant

Further the bandwidth parameter K in the kernel estimates is chosen proportional to $c_T T^b$ for some $b \in (1/4, 2/3)$ where c_T is slowly varying at infinity (i.e. $c_{Tx}/c_T \to 1, \forall x > 0$). $\square$

Analogously to the RRR estimator derived from the OLS estimator we derive the new fully modified RRR estimator (henceforth denoted as `FM-RRR`) from the FM-estimator using the SVD

$$\begin{aligned} \tilde{\Xi}_+ \hat{\beta}^+_{OLS,r} \langle z_t, z_t \rangle^{1/2} &= \hat{U}^+_n \hat{S}^+_n (\hat{V}^+_n)' + \hat{R}^+_n, \\ \tilde{\Xi}_+ &= \left(\langle y^\pi_t, y^\pi_t \rangle - \hat{\Omega}_{\tilde{u},\Delta z} \hat{\Omega}^{-1}_{\Delta z,\Delta z} (\langle \Delta z_t, y^\pi_t \rangle - \hat{\Delta}_{\Delta z,\Delta y^\pi}) - (\langle y^\pi_t, \Delta z_t \rangle - \hat{\Delta}_{\Delta y^\pi,\Delta z}) \hat{\Omega}^{-1}_{\Delta z,\Delta z} \hat{\Omega}_{\Delta z,\hat{u}} \right)^{-1/2} \end{aligned} \tag{4}$$

where as before $\hat{U}^+_n$ denotes the matrix of left singular vectors, $\hat{S}^+_n = \mathrm{diag}(\hat{s}^+_1, \hat{s}^+_2, \ldots, \hat{s}^+_n)$ is the diagonal matrix containing the dominant estimated singular values $\hat{s}^+_1 \geq \hat{s}^+_2 \geq \ldots \geq \hat{s}^+_n > 0$ decreasing in size and the columns of $\hat{V}^+_n$ contain the corresponding right singular vectors.

The estimator under the rank restriction $\mathrm{rank}(\beta) = n$ then is defined as

$$\hat{\beta}^+_{RRR,r} = (\tilde{\Xi}_+)^{-1} \hat{U}^+_n \hat{S}^+_n (\hat{V}^+_n)' \langle z_t, z_t \rangle^{-1/2}, \hat{\beta}^+_{RRR,u} = \hat{\beta}^+_{OLS,u} - (\hat{\beta}^+_{OLS,r} - \hat{\beta}^+_{RRR,r}) \langle z^r_t, z^u_t \rangle \langle z^u_t, z^u_t \rangle^{-1}. \tag{5}$$

3 Results

In this paper the following assumptions on the data generating process (dgp) will be used:
Assumption P: The process $(y_t)_{t\in\mathbb{Z}}$ is generated according to (1) with $u_t = \Lambda \varepsilon_t$ ($\Lambda \in \mathbb{R}^{s\times k}$ of full row rank) where $(z^r_t)_{t\in\mathbb{Z}}$ and $(z^u_t)_{t\in\mathbb{Z}}$ are processes such that for some orthogonal matrices $H_r = [H_{r,\|}, H_{r,\perp}], H_u = [H_{u,\|}, H_{u,\perp}]$ ($H_{r,\|} \in \mathbb{R}^{m_r \times c_r}, H_{u,\|} \in \mathbb{R}^{m_u \times c_u}$) we have

$$\mathrm{diag}(\Delta(L) I_{c_r}, I_{m_r - c_r}) H'_r z^r_t = v_t, \quad t \in \mathbb{Z}, \quad \mathrm{diag}(\Delta(L) I_{c_u}, I_{m_u - c_u}) H'_u z^u_t = w_t, \quad t \in \mathbb{Z},$$

where $\Delta(L) = 1 - L$ denotes the difference operator (L denoting the backward shift operator) and the joint vector $\nu_t := [v'_t, w'_t]'$ is a stationary process generated according to

$$\nu_t = \sum_{j=1}^{\infty} C_j \varepsilon_{t-j}$$

where $\sum_{j=1}^{\infty} j^a \|C_j\| < \infty$ for some $a > 3/2$ and where for the transfer function $c(z) := [c_v(z)', c_w(z)']' = \sum_{j=1}^{\infty} C_j z^j$ (with z denoting a complex variable) the matrix $c(1)$ is of full

row rank. Additionally it is assumed that

$$\mathbb{E}\left[\begin{array}{c} H'_{r,\perp}z_t^r \\ H'_{u,\perp}z_t^u \end{array}\right]\left[\begin{array}{c} H'_{r,\perp}z_t^r \\ H'_{u,\perp}z_t^u \end{array}\right]' > 0.$$

Here $(\varepsilon_t)_{t\in\mathbb{Z}}$ is an iid process with zero mean, nonsingular variance Σ and finite fourth moments. Finally $H'_{r,\|}z_0^r = 0$ and $H'_{u,\|}z_0^u = 0$. $\square$

Note that summation for ν_t starts at $j = 1$. Thus uncorrelatedness of the regressors with the noise is built into the assumptions. The assumptions imply that z_t^r and z_t^u are I(1) processes such that the cointegrating rank of the joint process equals the sum of the cointegrating ranks of the two processes.

The assumption of zero initial conditions is not important and can be replaced with the assumption of deterministic initial conditions, i.e. assuming that modeling is performed conditional on initial conditions.

The noise is assumed to constitute an iid sequence which is somewhat restrictive. Weaker assumptions are possible but make the asymptotic distributions more involved. Further note that the same noise ε_t is used to generate the regressors as well as the residuals in the estimation equation. Consequently lagged y_t's are admitted as regressors and some dynamics may be included in the model, alleviating the iid assumption.

Furthermore these assumptions exclude deterministic terms such as the constant as regressors which are discussed separately below.

The assumptions on the data generating process lead to the following representation result:

Theorem 3.1 *Let Assumption P hold where $n = rank(b_r), b = [b_r, b_u]$.*

(I) Let $c_y \leq n$ denote the rank of $b_r H_{r,\|}$. Then the cointegrating rank of $(z_t^r)_{t\in\mathbb{Z}}$ is $m_r - c_r$ and the cointegrating rank of $(y_t - b_u z_t^u)_{t\in\mathbb{Z}}$ is $s - c_y$.

(II) There exist nonsingular matrices $\mathcal{T}_y \in \mathbb{R}^{s\times s}, \mathcal{T}_{z,r} \in \mathbb{R}^{m_r\times m_r}$ and $\mathcal{T}_{z,u} \in \mathbb{R}^{m_u\times m_u}$ such that

$$\tilde{y}_t = \left[\begin{array}{c} \tilde{y}_{t,1} \\ \tilde{y}_{t,2} \end{array}\right] = \mathcal{T}_y(y_t - b_u z_t^u) = \tilde{b}_r \tilde{z}_t + \tilde{\varepsilon}_t = \left[\begin{array}{ccc} I_{c_y} & 0 & 0 \\ 0 & 0 & \tilde{b}_{2,3} \end{array}\right]\left[\begin{array}{c} \tilde{z}_{t,1} \\ \tilde{z}_{t,2} \\ \tilde{z}_{t,3} \end{array}\right] + \left[\begin{array}{c} \tilde{\varepsilon}_{t,1} \\ \tilde{\varepsilon}_{t,2} \end{array}\right],$$

$$\tilde{z}_t = \mathcal{T}_{z,r} z_t^r = \left[\begin{array}{c} \tilde{z}_{t,1} \\ \tilde{z}_{t,2} \\ \tilde{z}_{t,3} \end{array}\right], \tilde{z}_t^u = \mathcal{T}_{z,u} z_t^u = \left[\begin{array}{c} \tilde{z}_{t,1}^u \\ \tilde{z}_{t,2}^u \end{array}\right]$$

where $\Delta(L)\tilde{z}_{t,1} = \tilde{c}_{z,1}(L)\varepsilon_t, \Delta(L)\tilde{z}_{t,2} = \tilde{c}_{z,2}(L)\varepsilon_t, \Delta(L)\tilde{z}_{t,1}^u = \tilde{c}_{z,u}(L)\varepsilon_t, t \in \mathbb{Z}$ (L denoting the backward shift operator) and the matrix $[\tilde{c}_{z,1}(1)', \tilde{c}_{z,2}(1)', \tilde{c}_{z,u}(1)']$ is of full column rank, and $(\tilde{z}_{t,3})_{t\in\mathbb{Z}}$ and $(\tilde{z}_{t,2}^u)_{t\in\mathbb{Z}}$ are stationary processes with nonsingular spectrum at $z = 1$.

The result is proved in Appendix A. It builds the main representation of the regression on which the asymptotic results are based upon. Note that the matrices $\mathcal{T}_y, \mathcal{T}_{z,u}$ and $\mathcal{T}_{z,r}$ separating the non-stationary and stationary directions of the various processes are not unique and the theorem only ascertains the existence. The restrictions on the ranks of the various matrices ensures that the various components are either stationary processes which are not over differenced or integrated processes which are not cointegrated.

Under these assumptions it is well known that the OLS estimators are weakly consistent (Park and Phillips, 1988; Park and Phillips, 1989). Furthermore almost sure consistency as well as the convergence rate $\hat{\beta}_{OLS} - b = O(\sqrt{\log\log T/T})$ (i.e. $\sqrt{T/\log\log T}(\hat{\beta}_{OLS} - b)$ is almost surely (a.s.) bounded) can be derived, see e.g. (Bauer, 2009). Additionally their asymptotic distribution is also well documented: Let $\mathcal{T}_z = \operatorname{diag}(\mathcal{T}_{z,r}, \mathcal{T}_{z,u})$ and let $D_z = \operatorname{diag}(D_{z,r}, D_{z,u}), D_{z,r} = \operatorname{diag}(T^{-1}I_{c_r}, T^{-1/2}I_{m_r-c_r}), D_{z,u} = \operatorname{diag}(T^{-1}I_{c_u}, T^{-1/2}I_{m_u-c_u})$. Then one obtains

$$\mathcal{T}_y(\hat{\beta}_{OLS} - b)\mathcal{T}_z^{-1}D_z^{-1} \xrightarrow{d} \left[\begin{array}{cccc} M_r & Z_r & M_u - M_r N_r & Z_u - Z_r \mathbb{E}\tilde{z}_{t,3}(\tilde{z}_{t,2}^u)'(\mathbb{E}\tilde{z}_{t,2}^u(\tilde{z}_{t,2}^u)')^{-1} \end{array}\right]$$

where (using the notation[1] $f(E, W) = \int dEW'(\int WW')^{-1}$)

$$\begin{aligned}
M_r &= f(\mathcal{T}_y \Lambda W, W_z^{\Pi}), \\
M_u &= f(\mathcal{T}_y \Lambda W, W_u), \\
N_r &= \int W_z W_u' \left(\int W_u W_u'\right)^{-1}, \\
\operatorname{vec}\left[T^{-1/2}\sum_{t=1}^{T} \mathcal{T}_y \Lambda \varepsilon_t (\tilde{z}_{t,3}^{\pi})' \langle \tilde{z}_{t,3}^{\pi}, \tilde{z}_{t,3}^{\pi}\rangle^{-1}\right] &\xrightarrow{d} \operatorname{vec}(Z_r), \\
\operatorname{vec}\left[T^{-1/2}\sum_{t=1}^{T} \mathcal{T}_y \Lambda \varepsilon_t (\tilde{z}_{t,2}^{u})' \langle \tilde{z}_{t,2}^{u}, \tilde{z}_{t,2}^{u}\rangle^{-1}\right] &\xrightarrow{d} \operatorname{vec}(Z_u),
\end{aligned}$$

where $\tilde{z}_{t,3}^{\pi} := \tilde{z}_{t,3} - \langle \tilde{z}_{t,3}, \tilde{z}_{t,2}^u\rangle \langle \tilde{z}_{t,2}^u, \tilde{z}_{t,2}^u\rangle^{-1}\tilde{z}_{t,2}^u$. W denotes the Brownian motion corresponding to $(\varepsilon_t)_{t\in\mathbb{N}}$ and $W_z = \tilde{c}_{z,1:2}(1)W, W_u = \tilde{c}_{z,u}(1)W, W_z^{\Pi} = W_z - \int W_z W_u'(\int W_u W_u')^{-1}W_u$. Further $\operatorname{vec}(Z_r)$ and $\operatorname{vec}(Z_u)$ are normally distributed with mean zero (vec denotes columnwise vectorisation). Finally $\tilde{c}_{z,1:2}(1) := [\tilde{c}_{z,1}(1)', \tilde{c}_{z,2}(1)']'$.

The next theorem, which is the main contribution of this paper, extends these results to the RRR estimators:

[1] Here and below $\int dEW'$ is the usual shorthand notation for $\int_0^1 dE(w)W(w)'$ for Brownian motions $E(w), W(w), w \in [0, 1]$. Analogously $\int WW'$ is short for $\int_0^1 W(w)W(w)'dw$.

Theorem 3.2 *(I) Let the assumptions of Theorem 3.1 hold. Then* $\hat{\beta}_{RRR} - b = O((\log T)^6/\sqrt{T})$. *Furthermore let* $\hat{\beta}_{RRR,r}$ *and* $\hat{\beta}_{OLS,r}$ *denote the coefficients corresponding to* z_t^r. *Then the asymptotic distribution of* $\hat{\beta}_{RRR,r}$ *can be found from*

$$[\mathcal{T}_y(\hat{\beta}_{RRR,r} - \hat{\beta}_{OLS,r})\mathcal{T}_{z,r}^{-1}]D_{z,r}^{-1} \xrightarrow{d} -\left[\begin{bmatrix} -\Xi \\ I \end{bmatrix} (I - \tilde{O}_2\tilde{O}_2^\dagger) M_{r,2} \begin{bmatrix} -Y_{21}Y_{11}^{-1} & I \end{bmatrix} \begin{bmatrix} \tilde{R} \\ (I - \tilde{O}_2\tilde{O}_2^\dagger)\tilde{Z}_{r,2}P \end{bmatrix} \right]$$

where

$$\begin{aligned}
\Xi &= -[I,0]\mathcal{T}_y\Lambda\mathbb{E}\varepsilon_t\tilde{y}_{t,2}'(\mathbb{E}\tilde{y}_{t,2}^\Pi\tilde{y}_{t,2}')^{-1}, \\
M_{r,2} &= f([0,I]\mathcal{T}_y\Lambda W, W_{z,2}^\Pi - Y_{21}Y_{11}^{-1}W_{z,1}^\Pi), \\
Y_{i1} &= \int W_{z,i}^\Pi(W_{z,1}^\Pi)', i = 1, 2, \\
P &= I - \mathbb{E}\tilde{z}_{t,3}^\Pi(\tilde{z}_{t,3}^\Pi)'\Gamma_{3,2}\Gamma_{3,2}^\dagger, \Gamma_{3,2}^\dagger = (\Gamma_{3,2}'\mathbb{E}\tilde{z}_{t,3}^\Pi(\tilde{z}_{t,3}^\Pi)'\Gamma_{3,2})^{-1}\Gamma_{3,2}', \\
\tilde{b}_{2,3} &= \tilde{O}_2\Gamma_{3,2}', \tilde{O}_2^\dagger = (\tilde{O}_2'(\mathbb{E}\tilde{y}_{t,2}^\Pi\tilde{y}_{t,2}')^{-1}\tilde{O}_2)^{-1}\tilde{O}_2'(\mathbb{E}\tilde{y}_{t,2}^\Pi\tilde{y}_{t,2}')^{-1}.
\end{aligned}$$

Here $\tilde{Z}_{r,2} = [0,I]Z_r$, $W_z^\Pi = [(W_{z,1}^\Pi)', (W_{z,2}^\Pi)']'$ *and* $\tilde{z}_{t,3}^\Pi = \tilde{z}_{t,3} - \mathbb{E}\tilde{z}_{t,3}(\tilde{z}_{t,2}^u)'(\mathbb{E}\tilde{z}_{t,2}^u(\tilde{z}_{t,2}^u)')^{-1}\tilde{z}_{t,2}^u$ *and* $\tilde{y}_{t,2}^\Pi$ *is defined analogously. Finally* $\tilde{R}$ *is defined in Lemma A.9. Correspondingly letting* $\tilde{\beta}_{RRR,u}$ *and* $\tilde{\beta}_{OLS,u}$ *denote the coefficients corresponding to* z_t^u *then*

$$\mathcal{T}_y(\hat{\beta}_{RRR,u} - \hat{\beta}_{OLS,u})\mathcal{T}_{z,u}^{-1}D_{z,u}^{-1} = -\mathcal{T}_y(\hat{\beta}_{RRR,z} - \hat{\beta}_{OLS,z})\mathcal{T}_{z,r}^{-1}D_{z,u}^{-1} \begin{bmatrix} N_r & 0 \\ 0 & \mathbb{E}\tilde{z}_{t,3}(\tilde{z}_{t,2}^u)'(\mathbb{E}\tilde{z}_{t,2}^u(\tilde{z}_{t,2}^u)')^{-1} \end{bmatrix} + o_P(1).$$

(II) All results hold true in the situation that all observations are demeaned or detrended prior to estimation, if a.s. rates are replaced with in probability rates, the Brownian motions are replaced by their corresponding demeaned or detrended version and if additionally to the assumptions above the condition $\sum_{j=1}^\infty j^a\|C_j\| < \infty$ *holds for some* $a > 3$.

Note that the decomposition of $\tilde{b}_{2,3}$ is not specified. The asymptotic distribution does not depend on the actual choice.

The theorem shows how the inclusion of the rank constraint affects the estimation error which is given as a sum of the error for the unrestricted estimate plus a correction term. All coefficients corresponding to the nonstationary directions in z_t are estimated T-consistent and asymptotically the estimation errors have 'matrix unit root' distributions, whereas for directions in which $(z_t)_{t\in\mathbb{N}}$ is stationary the coefficients are only $\sqrt{T}$ consistent and the errors are asymptotically normal. The proof of this theorem is given in Appendix A. Note that the larger bounds in the almost sure convergence rates for the restricted estimator reflects

only the techniques of proof used and not the accuracy of the estimators which is more appropriately represented in the distributional results. I.e. the larger bounds for the rank restricted estimator mirrors our inability to prove the tighter bounds rather than the relative accuracy of the estimators.

In the fully modified case conditions for consistency and the asymptotic distribution of the unrestricted estimator is provided in (Phillips, 1995): Under the assumptions on the kernel provided in Assumptions K one obtains:

$$\mathcal{T}_y(\hat{\beta}^+_{OLS} - b)\mathcal{T}_z^{-1} D_z^{-1} \xrightarrow{d} \begin{bmatrix} M_r^+ & Z_r & M_u^+ - M_r^+ N_r & Z_u - Z_r \mathbb{E}\tilde{z}_{t,3}(\tilde{z}^u_{t,2})'(\mathbb{E}\tilde{z}^u_{t,2}(\tilde{z}^u_{t,2})')^{-1} \end{bmatrix}$$

where

$$\begin{aligned} B &= \Lambda W - \Omega^{:,n}_{u,\Delta z}(\Omega^{n,n}_{\Delta z,\Delta z})^{-1} \begin{bmatrix} \tilde{c}_{z,1:2}(1)W \\ \tilde{c}_{z,u}(1)W \end{bmatrix}, \\ M_r^+ &= f(B, W_z^{\Pi}), \\ M_u^+ &= f(B, \tilde{c}_{z,u}(1)W). \end{aligned}$$

Here the superscript n refers to the nonstationary directions in $[\tilde{z}_t', (\tilde{z}_t^u)']'$ and the matrices $\Omega^{:,n}_{u,\Delta z}$ and $\Omega^{n,n}_{\Delta z,\Delta z}$ are composed of the respective columns and rows corresponding to the nonstationary components.

The next theorem discusses the properties of the corresponding rank restricted estimator:

Theorem 3.3 *Let the assumptions of Theorem 3.2 hold and additionally assume that a kernel function fulfilling assumptions K is used in the nonparametric estimation of the long run variances. Let $\tilde{\beta}^+_{RRR}$ denote the FM-RRR estimator (based on the fully modified estimator $\tilde{\beta}^+_{OLS}$) defined in (5) for the weight $\tilde{\Xi}_+$ as defined in (4). Then using the notation of Theorem 3.2 it holds that*

$$[\mathcal{T}_y(\hat{\beta}^+_{RRR,r} - \hat{\beta}^+_{OLS,r})\mathcal{T}_{z,r}^{-1}]D_{z,r}^{-1} \xrightarrow{d} -\left[\begin{bmatrix} \Xi \\ I \end{bmatrix} (I - \tilde{O}_2\tilde{O}_2^{\dagger}) M_2^+ \begin{bmatrix} -Y_{21}Y_{11}^{-1} & I \end{bmatrix} \begin{bmatrix} \tilde{R} \\ (I - \tilde{O}_2\tilde{O}_2^{\dagger})\tilde{Z}_{r,2}P \end{bmatrix} \right].$$

where $M_2^+ = f([0, I]\mathcal{T}_y \Lambda B, W^{\Pi}_{z,2} - Y_{21}Y_{11}^{-1}W^{\Pi}_{z,1})$.

Therefore the relation between the restricted and the unrestricted regressions are identical for the conventional and the fully modified case. Also the expressions for the two sets of estimators are identical except for the use of W in the conventional case which is replaced by B in the fully modified case. Therefore it follows that the distribution in the direction of (asymptotically) stationary components of w_t is identical for both estimators. Hence in

the case that $c_z = 0$ and therefore no integration is present the conventional and the fully modified estimators have the same asymptotic distribution. This is true for the restricted and the unrestricted estimates. We refrain from a more complete discussion on the properties of the fully modified estimator since for the unrestricted case these are well documented in the literature. Instead a number of special cases will be discussed below.

4 Special Cases

First consider the case where all included variables are stationary. In that case $\mathcal{T}_y = I, \mathcal{T}_z = I$ can be used and the asymptotic distribution of the vectorizations of $\sqrt{T}(\hat{\beta}_{OLS} - b)$ and $\sqrt{T}(\hat{\beta}^+_{OLS} - b)$ are both normal with mean zero and variance $(\mathbb{E}z_t z_t')^{-1} \otimes \Lambda\Sigma\Lambda'$ which equals the distribution of

$$\mathrm{vec}\left(\left[Z_r, \quad Z_u - Z_r \mathbb{E}\tilde{z}_{t,3}(\tilde{z}^u_{t,2})'(\mathbb{E}\tilde{z}^u_{t,2}(\tilde{z}^u_{t,2})')^{-1}\right]\right)$$

noting that in this case $\tilde{z}_t = \tilde{z}_{t,3} = z^r_t, \tilde{z}^u_{t,2} = z^u_t$ can be chosen. The correction due to the rank restriction for β_r equals the vectorization of

$$-(I - \tilde{O}_2\tilde{O}_2^\dagger)Z_r(I - \mathbb{E}\tilde{z}_{t,3}(\tilde{z}_{t,3})'\Gamma_{32}\Gamma_{32}^\dagger).$$

The corresponding correction to b_u follows. On total hence one obtains as the asymptotic distribution of the RRR-estimator for b_r the distribution of $Z_r - (I - \tilde{O}_2\tilde{O}_2^\dagger)Z_r(I - \mathbb{E}\tilde{z}_{t,3}(\tilde{z}_{t,3})'\Gamma_{32}\Gamma_{32}^\dagger)$.

This asymptotic distribution (for a generic case) has previously been documented in (Reinsel and Velu, 1998) on p. 45 (2.36) albeit in a different form which is less accessible. On p. 46 a more explicit expression for the case $n = 1$ is given. It is straightforward to see that the expressions in this special case are identical while the formula provided above also provide insights in the general case. It must be noted, however, that these expressions are not new and have been used already e.g. in (Bauer, Deistler and Scherrer, 1999).

The consequences of the correction using the rank restriction are the following: Premultiplying the asymptotic distribution with $\tilde{O}_2^\dagger$ one notices that the rank restriction does not influence the distribution in these directions. In the orthogonal complement, however, the distribution is changed from $x'Z_r$ to $x'Z_r\mathbb{E}\tilde{z}_{t,3}(\tilde{z}_{t,3})'\Gamma_{32}\Gamma_{32}^\dagger$ and hence projected onto the rows corresponding to the space spanned by the columns of Γ_{32}. The analogous statements hold

for the postmultiplication with Γ_{32}. Note that these arguments also hold in the general case for the $(2,3)$ block of $\tilde{b}_r$.

As a second special case consider the VAR(1) I(1) model of Anderson (2002). For simplicity of notation the transformed system will be used which in the notation of Anderson (2002) is stated as

$$\Delta X_t = \Upsilon X_{t-1} + W_t = \begin{bmatrix} 0_{c_y \times c_y} & 0 \\ 0 & \Upsilon_{22} \end{bmatrix} X_{t-1} + W_t, t \in \mathbb{N} \tag{6}$$

for $X_0 = 0$ where Υ_{22} is nonsingular. This defines an I(1) process $X_t \in \mathbb{R}^s$ whose first component, $X_{t,1} \in \mathbb{R}^{c_y}$ say, is integrated, the remaining, $X_{t,2} \in \mathbb{R}^{s-c_y}$ say, being stationary for $|\lambda_{max}(I + \Upsilon_{22})| < 1$ which is assumed in the following. The variance of the iid white noise W_t is taken to be $[\Sigma_{ij}^W]_{i,j=1,2}$ which is partitioned according to the partitioning of X_t. In our notation no transformation matrices are needed since the system is already in the appropriate coordinate system. Hence $y_t = \tilde{y}_{t,2} = \Delta X_t$ is stationary, $\tilde{z}_{t,2} = X_{t-1,1}, \tilde{z}_{t,3} = X_{t-1,2}, z_t^u$ does not occur. Consequently $\tilde{b}_r = \Upsilon = [\Upsilon_{:,1}, \Upsilon_{:,2}]$ and b_u does not occur.

In this situation (Anderson, 2002) gives the asymptotic distribution of the unrestricted and the restricted estimates of Υ. Consider first $\Upsilon_{:,1}$, i.e. the first block column. Then Theorem 1 of (Anderson, 2002) states that $T\tilde{\Upsilon}_{:,1,OLS} \xrightarrow{d} J_{:,1} I_{11}^{-1}$ which in our notation equals $f(W, W_1)$ where W is the Brownian motion according to $u_t = W_t$ and W_1 denotes the corresponding first block. For $\tilde{\beta}_{RRR,r}$ (Anderson, 2002) states the asymptotic distribution of the first block column as

$$T\tilde{\Upsilon}_{:,1,RRR} \xrightarrow{d} \begin{bmatrix} 0 \\ J_{2.1} I_{11}^{-1} \end{bmatrix}, J_{2.1} = [-\Sigma_{WW}^{21}(\Sigma_{WW}^{11})^{-1}, I] J_{:,1}.$$

From Theorem 3.2 we obtain

$$T\tilde{\beta}_{:,1,RRR} = T\tilde{\beta}_{:,1,OLS} + T(\tilde{\beta}_{:,1,RRR} - \tilde{\beta}_{:,1,OLS}) \xrightarrow{d} f(W, W_1) - (I - \tilde{O}_2 \tilde{O}_2^\dagger) f(W, W_1) = \tilde{O}_2 \tilde{O}_2^\dagger J_{:,1} I_{11}^{-1}.$$

The second block column of $\tilde{b}_r$ provides the decomposition $\tilde{O}_2 := [0, I]', \Gamma_{32}' := \Upsilon_{22}$. Here $\tilde{O}_2 \tilde{O}_2^\dagger = [0, I]'([0, I]\Sigma_{y2,y2}^{-1}[0, I]')^{-1}[0, I]\Sigma_{y2,y2}^{-1}$ where

$$\begin{aligned} \Sigma_{y2,y2}^{-1} &= \begin{bmatrix} \Sigma_{WW}^{11} & \Sigma_{WW}^{12} \\ \Sigma_{WW}^{21} & Q \end{bmatrix}^{-1} \\ &= \begin{bmatrix} (\Sigma_{WW}^{11})^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -(\Sigma_{WW}^{11})^{-1}\Sigma_{WW}^{12} \\ I \end{bmatrix} \left(Q - \Sigma_{WW}^{21}(\Sigma_{WW}^{11})^{-1}\Sigma_{WW}^{12}\right)^{-1} [-\Sigma_{WW}^{21}(\Sigma_{WW}^{11})^{-1}, I] \end{aligned}$$

(for some matrix Q) according to the block matrix inversion. Thus $\tilde{O}_2 \tilde{O}_2^\dagger = [0, I]'[-\Sigma_{WW}^{21}(\Sigma_{WW}^{11})^{-1}, I]$ showing the identity of the expressions.

For the FM estimator note that the involved long run covariances equal

$$\Omega_{u,\Delta z}^{:,n} = \Sigma_{WW}^{:,1}, \quad \Omega_{\Delta z,\Delta z}^{n,n} = \Sigma_{WW}^{11}$$

with all other terms being zero. Correspondingly

$$\Omega_{u,\Delta z}^{:,n}(\Omega_{\Delta z,\Delta z}^{n,n})^{-1} = [I, (\Sigma_{WW}^{11})^{-1}\Sigma_{WW}^{12}]' \Rightarrow J_{2.1}I_{11}^{-1} = f(B,W).$$

Thus it follows that the coefficients to the nonstationary regressors for the FM-estimator have the same asymptotic distribution as the RRR estimators. Adding the rank restriction in this case does not change the *asymptotic distribution* while it might well influence the finite sample properties. It is straightforward to show that in this case also the RRR-FM estimator has the same distribution for the columns corresponding to the integrated regressors.

With respect to the stationary directions it is easy to see that $P = 0$ since $\Gamma_{32} = \Upsilon_{22}$ is invertible. Consequently the RRR estimator and the OLS estimator have the same asymptotic distribution in the columns corresponding to the stationary regressors. Since FM and OLS estimators have the same asymptotic behavior for stationary regressors all four estimators show the same asymptotic behavior in these columns. The underlying reason for this is that the rank restriction exclusively applies to the nonstationary restrictions where the corresponding coefficient is restricted to zero. For the stationary regressors there are no other rank restrictions in this case.

Adding additional lagged first differences to (6) the AR(p) setting with transformed coordinates is obtained. The additional coefficients are not restricted (except for the seldom imposed restriction of stability of the corresponding transfer function) and hence in this case $z_t^u = \tilde{z}_{t,2}^u = [\Delta X_{t-1}', \ldots, \Delta X_{t-p+1}']'$, i.e. additional stationary regressors are present. It is well known that in this case (using the usual notation such that $\Upsilon = \alpha\beta'$ where $\alpha_\perp'\alpha = 0, \beta_\perp'\beta = 0$ for orthogonal matrices $\alpha_\perp, \beta_\perp$ of maximal dimension such that the columns span the orthogonal complement of α, β respectively) we have

$$X_t \quad = \quad \beta_\perp(\Gamma_J)^{-1}\left(\alpha_\perp'\sum_{j=1}^{t-1}\varepsilon_{t-j} + X_1\right) + w_t, \tag{7}$$

for some stationary process w_t and nonsingular matrix Γ_J (expressions could be given but are not of importance in the following and hence omitted). In the example $\alpha' = [0, \Upsilon_{22}'], \beta' = [0, I]$ and thus $\alpha_\perp' = [I, 0], \beta_\perp = [I, 0]'$.

The changes in the asymptotic distribution are the following: W is unchanged while $W_z = \Gamma_J^{-1} W_1$. The stationary components change accordingly. Imposing the rank restriction (as is done in the Johansen quasi-ML estimators) does not change the asymptotic distribution of the coefficients corresponding to the stationary terms as in the AR(1) case presented above since Γ_{32} again is nonsingular and hence $P = 0$. Thus we obtain the same asymptotic distribution as in the non restricted case. This asymptotic distribution is also given in Theorem 13.5. of (Johansen, 1995).

For the coefficients corresponding to nonstationary coordinates we obtain analogously to above

$$T\tilde{\beta}_{:,1,RRR} = T\tilde{\beta}_{:,1,OLS} + T(\tilde{\beta}_{:,1,RRR} - \tilde{\beta}_{:,1,OLS}) \xrightarrow{d} \tilde{O}_2 \tilde{O}_2^{\dagger} f(W, W_z) = \begin{bmatrix} 0 \\ I \end{bmatrix} [-\Sigma_{WW}^{21}(\Sigma_{WW}^{11})^{-1}, I] f(W, W_z).$$

For the FM estimator note that the involved long run covariances equal

$$\Omega_{u,\Delta z}^{:,n} = \Sigma_{WW}^{:,1}(\Gamma_J^{-1})', \Omega_{\Delta z,\Delta z}^{n,n} = \Gamma_J^{-1}\Sigma_{WW}^{11}(\Gamma_J^{-1})'$$

due to the change in the nonstationary directions. Then $\Omega_{u,\Delta z}^{:,n}(\Omega_{\Delta z,\Delta z}^{n,n})^{-1} = [I, (\Sigma_{WW}^{11})^{-1}\Sigma_{WW}^{12}]'(\Gamma_J^{-1})'$ as above implies that again the unrestricted FM estimator has the same asymptotic distribution as the RRR estimator. This is remarkable since the FM estimator does not require the specification of the rank restriction. This has already been observed in (Phillips, 1995) but apparently did not draw the attention of the community.

5 Conclusions

In this paper the asymptotic properties for two estimators in a regression setting explicitly imposing a rank restriction are discussed. Beside providing (almost sure) rates of convergence also explicit expressions for the asymptotic distribution of transformed estimators (such that stationary and nonstationary coordinates are separated) are provided. These expressions reveal the main characteristics of the estimators and allow insights into the relative merits of the various methods such as the gain in asymptotic accuracy obtained by imposing the rank restriction. In particular it is shown that the fully modified estimators in many situations achieve the same asymptotic distribution as the rank restricted regression OLS estimators without imposing the rank restriction. This is an attractive feature in situations where the rank is not known.

The results contain a number of well known situations as special cases and even in some of these cases allow new insights as the previously published expressions for the asymptotic distribution are much more complicated to interpret.

Finally it must be noted that the results in this paper are seen to be intermediate results that might in many cases not seem to be relevant as they relate to transformed estimators where the transformations are not known during the estimation. Nevertheless, the results are important ingredients to explore the properties of procedures that use the RRR as an intermediate step. An important example are subspace methods in the case of cointegrated processes. These results will be presented elsewhere.

References

Anderson, T. W.: 2002, Reduced rank regression in cointegrated models, *Journal of Econometrics* **106**, 203–216.

Bauer, D.: 2009, An almost sure bound for the inverse of sample second moments of integrated processes, *Econometric Theory* **25**, 571–582.

Bauer, D., Deistler, M. and Scherrer, W.: 1999, Consistency and asymptotic normality of some subspace algorithms for systems without observed inputs, *Automatica* **35**, 1243–1254.

Bauer, D. and Wagner, M.: 2002, Estimating cointegrated systems using subspace algorithms, *Journal of Econometrics* **111**, 47–84.

Chatelin, F.: 1983, *Spectral Approximation of Linear Operators*, Academic Press.

Davidson, J.: 1994, *Stochastic Limit Theory*, Oxford University Press.

Hannan, E. J. and Deistler, M.: 1988, *The Statistical Theory of Linear Systems*, John Wiley, New York.

Johansen, S.: 1995, *Likelihood-Based Inference in Cointegrated Vector Auto-Regressive Models*, Oxford University Press, Oxford, UK.

Lai, T. L. and Wei, C. Z.: 1983, Asymptotic properties of general autoregressive models and strong consistency of least squares estimates of their parameters, *Journal of Multivariate Analysis* **13**, 1–23.

Lai, T. and Wei, C.: 1982, Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems, *Annals of Statistic* **10**(1), 154–166.

Larimore, W.: 1983, System identification, reduced order filters and modeling via canonical variate analysis, *in* H. S. Rao and P. Dorato (eds), *Proc. 1983 Amer. Control Conference 2.*, Piscataway, NJ, pp. 445–451.

Park, J. and Phillips, P.: 1988, Statistical inference in regressions with integrated processes: Part I, *Econometric Theory* **4**, 468–497.

Park, J. and Phillips, P.: 1989, Statistical inference in regressions with integrated processes: Part II, *Econometric Theory* **5**, 95–131.

Phillips, P.: 1995, Fully modified least squares and vector autoregression, *Econometrica* **63**(5), 1023–1078.

Phillips, P. and Solo, V.: 1992, Asymptotic for linear processes, *The Annals of Statistics* **20**, 971–1001.

Reinsel, G. and Velu, R.: 1998, *Multivariate Reduced-Rank Regression*, Springer, New York.

Sims, C., Stock, J. and Watson, M.: 1990, Inference in linear time series models with some unit roots, *Econometrica* **58**, 113–144.

A Proofs

Throughout the appendix the following notation will be heavily used: For a sequence of random matrices F_T with elements $F_{i,j,T}$ and a sequence of scalars g_T we will use the notation $F_T = o(g_T)$ if $\limsup_{T\to\infty} \max_{i,j} |F_{i,j,T}/g_T| \to 0$ almost surely (a.s.). Similarly $F_T = O(g_T)$ if there exists a constant M such that $\limsup_{T\to\infty} \max_{i,j} |F_{i,j,T}/g_T| \leq M$ a.s. The corresponding in probability versions are: $F_T = o_P(g_T)$ if $\max_{i,j} |F_{i,j,T}/g_T| \to 0$ in probability and $F_T = O_P(g_T)$ if for each $\varepsilon > 0$ there exists a constant $M(\varepsilon) < \infty$ such that $\lim_{T\to\infty} \mathbb{P}\{\max_{i,j} |F_{i,j,T}/g_T| > M(\varepsilon)\} < \varepsilon$. In all these statements T denotes the sample size. Therefore in particular convergence in distribution to a finite dimensional almost surely finite random variable implies the rate $O_P(1)$. Throughout convergence in probability will be denoted as $\stackrel{p}{\to}$ and convergence in distribution as $\stackrel{d}{\to}$. Almost sure (a.s.) convergence

is denoted as $\to$. $\|.\|$ denotes the Euclidean norm if not stated explicitly otherwise. $\|.\|_{Fr}$ is used to denote the Frobenius norm. As usual the integral $\int W_1 W_2'$ is short notation for $\int_0^1 W_1(\omega)W_2(\omega)'d\omega$ and $\int dW_1 W_2'$ is short for $\int_0^1 dW_1(\omega)W_2(\omega)'$. Here $W_1(\omega)$ and $W_2(\omega)$ are two Brownian motions on $[0,1]$.

A.1 Preliminary lemmas

Lemma A.1 *(I) Let $(\varepsilon_t)_{t\in\mathbb{Z}}$ denote a white noise sequence which fulfills the noise assumptions contained in Assumption P. Define $x_{t,1} := \sum_{j=1}^{t-1}\varepsilon_j, t \geq 2, x_1 = 0, v_t := c_v(L)\varepsilon_t := \sum_{i=1}^{\infty} C_{v,i}\varepsilon_{t-i}, t \in \mathbb{N}$, for some transfer function $c_v(z) := \sum_{j=1}^{\infty} C_{v,j}z^j$ where $\sum_{j=1}^{\infty}\|C_{v,j}\|j^a < \infty$ for some $a > 3/2$. Furthermore $n_t := \sum_{j=0}^{t-1} C_{n,j}x_{t-j,1}, t \in \mathbb{N}$, for a sequence $C_{n,j}$ such that $\sum_{j=0}^{\infty}\|C_{n,j}\|j^a < \infty, a > 3/2$ and for $c_n(z) = \sum_{j=0}^{\infty} C_{n,j}z^j$ it holds that $\det c_n(1) \neq 0$. Finally let $Q_T := \sqrt{\log\log T/T}$. Then*

$$
\begin{aligned}
\|\langle v_t, \varepsilon_t\rangle\| = O(Q_T) \quad &, \quad \|\langle v_t, v_t\rangle - \mathbb{E}v_t v_t'\| = O(Q_T), \\
\langle x_{t,1}, x_{t,1}\rangle = O(T\log\log T) \quad &, \quad \langle x_{t,1}, \varepsilon_t\rangle = O(\log T), \\
\|\langle x_{t,1}, v_t\rangle\| = O(\log T) \quad &, \quad \langle x_{t,1}, x_{t,1}\rangle^{-1} = O(Q_T^2).
\end{aligned}
$$

All expressions remain true if $x_{t,1}$ is replaced by n_t.

(II) Furthermore using $\Delta_{v,\Delta n} = \sum_{j=0}^{\infty}\mathbb{E}v_j(\Delta n_0)'$ where $\Delta n_t = c_n(L)\varepsilon_t, t \in \mathbb{Z}$ we have:

$$
\begin{aligned}
\langle v_t, n_t\rangle &\xrightarrow{d} \int c_v(1)dWW'c_n(1)' + \Delta_{v,\Delta n}, \\
T^{-1}\langle n_t, n_t\rangle &\xrightarrow{d} c_n(1)\int WW'c_n(1)', \\
vec(T^{1/2}\langle\varepsilon_t, v_t\rangle) &\xrightarrow{d} \mathcal{N}(0, \mathbb{E}v_t v_t' \otimes \mathbb{E}\varepsilon_t\varepsilon_t')
\end{aligned}
$$

where vec denotes column wise vectorization, $W(w)$ denotes the limiting Brownian motion corresponding to $T^{-1/2}\sum_{j=1}^{\lfloor wT\rfloor}\varepsilon_j$. Finally $\mathcal{N}(0,V)$ denotes a Gaussian random variable with mean zero and variance V.

(III) Let $x_t := [x_{t,\bullet}', x_{t,1}']'$ where $(x_{t,\bullet})_{t\in\mathbb{N}}$ fulfills the same restrictions as $(v_t)_{t\in\mathbb{N}}$ under (I) and $(x_{t,1})_{t\in\mathbb{N}}$ and $(n_t)_{t\in\mathbb{N}}$ are integrated and of the same form as $(n_t)_{t\in\mathbb{N}}$ under (I). Further let $(v_t)_{t\in\mathbb{N}}$ and $(w_t)_{t\in\mathbb{N}}$ be two stationary processes fulfilling the assumption of $(v_t)_{t\in\mathbb{N}}$ under (I). Let $(\varepsilon_t)_{t\in\mathbb{Z}}$ be as under (I). Let $^{\pi}$ denote the residuals of a regression onto x_t and let $^{\Pi}$ denote the corresponding limits (whenever the symbol is used the limit exists). Hence e.g. $v_t^{\pi} = v_t - \langle v_t, x_t\rangle\langle x_t, x_t\rangle^{-1}x_t$. Then

$$
\langle\varepsilon_t, v_t^{\pi}\rangle = \langle\varepsilon_t, v_t^{\Pi}\rangle + o(T^{-1/2}) = O(Q_T),
$$

$$
\begin{aligned}
\langle \varepsilon_t, n_t^\pi \rangle &= \langle \varepsilon_t, n_t \rangle - \langle \varepsilon_t, x_{t,1}\rangle \langle x_{t,1}, x_{t,1}\rangle^{-1} \langle x_{t,1}, n_t\rangle + o(1) = O(\log T (\log\log T)^2),\\
\langle v_t^\pi, w_t^\pi \rangle &= \langle v_t^\Pi, w_t^\Pi \rangle + O(Q_T) = O(1),\\
\langle v_t^\pi, n_t^\pi \rangle &= O(\log T(\log\log T)^2),\\
\langle n_t^\pi, n_t^\pi \rangle &= \langle n_t, n_t\rangle - \langle n_t, x_{t,1}\rangle \langle x_{t,1}, x_{t,1}\rangle^{-1}\langle x_{t,1}, n_t\rangle + o(T) = O(T(\log\log T)).
\end{aligned}
$$

(IV) Let $d_t := [1,t]'$ and $D_d := diag(1, T^{-1})$. For any process $(a_t)_{t\in\mathbb{N}}$ let $\bar{a}_t$ denote the detrended process $\bar{a}_t := a_t - \langle a_t, d_t\rangle \langle d_t, d_t\rangle^{-1} d_t$. Let $v_t = \sum_{j=0}^\infty C_{v,j}\varepsilon_{t-j}$ where $\sum_{j=0}^\infty j^2 \|C_{v,j}\| < \infty$. Then $\|\langle \bar{v}_t, \bar{v}_t\rangle - \mathbb{E} v_t v_t'\| = O_P(Q_T)$. Further for $(x_t)_{t\in\mathbb{N}}$ as in (I) it follows that $\langle \bar\varepsilon_t, \bar{x}_t\rangle = O_P(1), \langle \bar{v}_t, \bar{x}_t\rangle = O_P(1)$. The same holds for replacing x_t with $n_t := \sum_{j=0}^{t-1} C_{n,j} x_{t-j}$ if $\sum_{j=0}^\infty \|C_{n,j}\| j^3 < \infty$.

The following limit theorems hold:

$$
\langle \bar\varepsilon_t, \bar{x}_t\rangle \xrightarrow{d} \int d\bar{W}\bar{W}' \quad , \quad T^{-1}\langle \bar{x}_t, \bar{x}_t\rangle \xrightarrow{d} \int \bar{W}\bar{W}'
$$

where $\bar{W} := W - (\int W(s)ds)(4-6\omega) - (\int sW(s)ds)(12\omega - 6)$ denotes the demeaned and detrended Brownian motion associated with $(\varepsilon_t)_{t\in\mathbb{N}}$.
The analogous results also holds for the demeaned series $\bar{a}_t := a_t - \langle a_t, 1\rangle$ where $\bar{W} := W - (\int W(s)ds)$ appears in the asymptotic distributions.

Proof: (I) $\langle v_t, \varepsilon_t\rangle = O(Q_T)$ and $\langle v_t, v_t\rangle - \mathbb{E} v_t v_t' = O(Q_T)$ follow from Theorem 7.4.3. of Hannan and Deistler (1988). $\langle x_t, x_t\rangle = O(T\log\log T)$ follows from Theorem 3 of Lai and Wei (1983), $\langle x_t, \varepsilon_t\rangle = o(\log T)$ from Corollary 2 of Lai and Wei (1982). Both results only deal with the univariate case but the extension to the multivariate situation is obvious. This result also implies $\langle x_t, v_t\rangle = O(\log T)$ (see (Bauer, 2009), Lemma 4). The same result applies for n_t in place of x_t by splitting $n_t = c_n(1)x_t + n_t^*$ (Beveridge-Nelson decomposition, see e.g. Phillips and Solo, 1992) where

$$
\begin{aligned}
n_t^* = n_t - c_n(1)x_t &= \sum_{j=0}^{t-1} C_{n,j}x_{t-j} - \sum_{j=0}^{\infty} C_{n,j}x_t = (C_{n,0} - c_n(1))x_t + \sum_{j=1}^{t-1} C_{n,j}x_{t-j}\\
&= (C_{n,0} - c_n(1))\varepsilon_t + (C_{n,0} - c_n(1))x_{t-1} + \sum_{j=1}^{t-1} C_{n,j}x_{t-j}\\
&= (C_{n,0} - c_n(1))\varepsilon_t + (C_{n,0} + C_{n,1} - c_n(1))(\varepsilon_{t-1} + x_{t-2}) + \sum_{j=2}^{t-1} C_{n,j}x_{t-j}\\
&= \sum_{j=0}^{t-1} C^*_{n,j}\varepsilon_{t-j}
\end{aligned}
$$

where $C^*_{n,i} := -\sum_{j=i+1}^{\infty} C_{n,j}$. Due to the summability assumptions on $C_{n,j}$ the transfer function $c^*_n(z) := \sum_{i=0}^{\infty} C^*_{n,i} z^i$ fulfills the properties of Theorem 7.4.3. of Hannan and Deistler (1988). The result then follows from the assumed non-singularity of $c_n(1)$.

The univariate version of $\langle x_t, x_t\rangle^{-1} = O(Q_T^2)$ is contained in Lai and Wei (1982, p. 163). The multivariate version is showed in Bauer (2009).

(II) Since $\sum_{t=1}^{\lfloor wT \rfloor} \varepsilon_t/\sqrt{T} \Rightarrow W(w)$ (Davidson, 1994, Theorem 27.17) the convergence of $\langle v_t, n_t\rangle$ is e.g. given in Park and Phillips (1988, Lemma 2.1. (e)). The result for $T^{-1}\langle n_t, n_t\rangle$ is stated in part (c) of the same lemma. The central limit theorem is standard (cf. e.g. Hannan and Deistler, 1988, Lemma 4.3.4.) since $(\varepsilon_t)_{t\in\mathbb{N}}$ is an ergodic square integrable martingale difference sequence.

(III) The proof is based on the block matrix inversion formula

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -A^{-1}B \\ I \end{bmatrix} (D - CA^{-1}B)^{-1} \begin{bmatrix} -CA^{-1}, I \end{bmatrix} \tag{8}$$

applied to $\langle x_t, x_t\rangle$. As an example consider

$$\begin{aligned} \langle \varepsilon_t, v_t^\pi\rangle &= \langle \varepsilon_t, v_t\rangle - \langle \varepsilon_t, x_t\rangle\langle x_t, x_t\rangle^{-1}\langle x_t, v_t\rangle \\ &= \langle \varepsilon_t, v_t\rangle - \langle \varepsilon_t, x_{t,\bullet}\rangle\langle x_{t,\bullet}, x_{t,\bullet}\rangle^{-1}\langle x_{t,\bullet}, v_t\rangle - \langle \varepsilon_t, x_{t,\neg 1}\rangle\langle x_{t,\neg 1}, x_{t,\neg 1}\rangle^{-1}\langle x_{t,\neg 1}, v_t\rangle \end{aligned}$$

where $x_{t,\neg 1} := x_{t,1} - \langle x_{t,1}, x_{t,\bullet}\rangle\langle x_{t,\bullet}, x_{t,\bullet}\rangle^{-1} x_{t,\bullet}$. Therefore $\langle x_{t,\neg 1}, x_{t,\neg 1}\rangle = \langle x_{t,1}, x_{t,1}\rangle - O(\log T)O(1)O(\log T)$ and $\langle x_{t,\neg 1}, v_t\rangle = \langle x_{t,1}, v_t\rangle - O(\log T) = O(\log T)$ by (I). Thus

$$\begin{aligned} \langle \varepsilon_t, v_t^\pi\rangle &= \langle \varepsilon_t, v_t\rangle - \langle \varepsilon_t, x_{t,\bullet}\rangle\langle x_{t,\bullet}, x_{t,\bullet}\rangle^{-1}\langle x_{t,\bullet}, v_t\rangle + o((\log T)^3/T) \\ &= \langle \varepsilon_t, v_t\rangle - \langle \varepsilon_t, x_{t,\bullet}\rangle(\mathbb{E}x_{t,\bullet}x_{t,\bullet}')^{-1}\mathbb{E}x_{t,\bullet}v_t' + o(T^{-1/2}) \end{aligned}$$

since $\langle \varepsilon_t, x_{t,\bullet}\rangle = O(Q_T)$ and $(\mathbb{E}x_{t,\bullet}x_{t,\bullet}')^{-1}\mathbb{E}x_{t,\bullet}v_t' - \langle x_{t,\bullet}, x_{t,\bullet}\rangle^{-1}\langle x_{t,\bullet}, v_t\rangle = O(Q_T)$ as required. For $\langle v_t^\pi, w_t^\pi\rangle = \langle v_t, w_t^\pi\rangle$ the same arguments apply with the exception that now $\langle v_t, w_t\rangle = O(1)$ rather than $O(Q_T)$ leading to the second claim. The other claims follow in a similar manner from the bounds achieved under (I).

(IV) Only the results for detrending are shown, the analogous statements for the demeaned series are obvious from the given results. The derivations here use Lemma 1, p. 121 of Sims, Stock and Watson (1990). Lemma 1 (g) shows that $\langle v_t, d_t\rangle D_d = O_P(T^{-1/2})$ for stationary $(v_t)_{t\in\mathbb{N}}$ and $\langle x_t, d_t\rangle D_d = O_P(T^{1/2})$ for integrated $(x_t)_{t\in\mathbb{N}}$.

For $\langle \bar\varepsilon_t, \bar x_t\rangle = \langle \varepsilon_t, x_t\rangle - \langle \varepsilon_t, d_t\rangle\langle d_t, d_t\rangle^{-1}\langle d_t, x_t\rangle$ note that the first term converges in distribution according to (II). For the second term note that $\sqrt{T}\langle \varepsilon_t, d_t\rangle D_d \xrightarrow{d} [\int dW, \int \omega dW]$, $D_d\langle d_t, d_t\rangle D_d$ converges to a constant nonsingular matrix and $T^{-1/2}D_d\langle d_t, x_t\rangle \xrightarrow{d} [\int W, \int \omega W]'$. The last two statements follow from Lemma 1 (a) and (c) of Sims et al. (1990). Therefore

$$\langle \bar\varepsilon_t, \bar x_t\rangle \xrightarrow{d} \int dWW' - \left[\int dW, \int \omega dW\right] \begin{bmatrix} \int 1 & \int \omega \\ \int \omega & \int \omega^2 \end{bmatrix}^{-1} \begin{bmatrix} \int W' \\ \int \omega W' \end{bmatrix}.$$

This shows that the Brownian motion in the limiting expression is demeaned and detrended.

If ε_t is replaced by v_t convergence in distribution still holds, but the limits change.

The evaluations for $T^{-1}\langle \bar{x}_t, \bar{x}_t\rangle$ follow the same lines and are omitted.

Decomposing $n_t = c_n(1)x_t + n_t^*$ as above shows that in the above calculations x_t can be replaced with n_t without changing the orders of convergence.

Finally if the time trend is omitted and only demeaning is performed the results can be shown analogously using the arguments given above. $\square$

Lemma A.2 *Under the assumptions of Theorem 3.3 the following holds true:*

$$\begin{aligned}
\hat{\Omega}_{\tilde{u},\Delta z}\hat{\Omega}^{-1}_{\Delta z,\Delta z} &= \begin{bmatrix} \Omega^{:,n}_{u,\Delta z}(\Omega^{n,n}_{\Delta z,\Delta z})^{-1} + o_P(1) & O_P(1) \end{bmatrix}, \\
\langle \Delta \check{z}_t, \check{z}_t\rangle - \hat{\Delta}_{\Delta z,\Delta z} &= \begin{bmatrix} \int dB[\begin{bmatrix} \tilde{c}_{z,1:2}(1) \\ \tilde{c}_{z,u}(1) \end{bmatrix} W]' + o_P(1) & O_P(K^{-2}) + O_P(1/\sqrt{KT}) \\ T^{-1}\check{z}_{T,2}\check{z}'_{T,1} + O_P(K^{-2}) + O_P(1/\sqrt{TK}) & O_P(K^{-2}) \end{bmatrix}, \\
\hat{\Delta}_{\tilde{u},\Delta z} &= \begin{bmatrix} \Delta^{:,n}_{u,\Delta z} + O_P((K/T)^{1/2}) & O_P(1/\sqrt{KT}) \end{bmatrix}, \\
(\langle \Delta \check{z}_t, \tilde{u}_t\rangle - \tilde{\Delta}_{\Delta z,\Delta u}) &= \begin{bmatrix} O_P(K^{-2}) + O_P(1/\sqrt{KT}) \\ O_P(K^{-2}) \end{bmatrix}
\end{aligned}$$

where $B =$

Here the notation refers to the transformed vectors $\check{z}_t = [(\tilde{z}_{t,1})', (\tilde{z}_{t,2})', (\tilde{z}^u_{t,1})', (\tilde{z}_{t,3})', (\tilde{z}^u_{t,2})']'$ *where the nonstationary components of both vectors (first block row) and the stationary components (second block row) of these vectors are separated. K denotes the kernel bandwidth parameter (see Assumptions K). Furthermore* $\Delta_{w,\Delta v} = \mathbb{E}w_t v_t'$ *for stationary processes* $(v_t)_{t\in\mathbb{Z}}$ *and* $(w_t)_{t\in\mathbb{Z}}$.

The proof of all but the last of these facts can be found in Phillips (1995, Lemma 8.1). The last fact can be easily derived from the infinite sum representation of $\Delta_{z,\Delta v}$.

Lemma A.3 *Let* $b_r = OG'$ *where* $G'S_p = I_n$. *Here* $S_p \in \mathbb{R}^{m_r\times n}$ *denotes a selector matrix (i.e. a matrix composed of n columns of I_{m_r}). Let $\tilde{\beta}_r$ denote an estimator of b_r such that* $\|\tilde{\beta}_r - b_r\|_{Fr} = o(a_T)$.

Assume that $\|\tilde{\Xi}_+ - \Xi_+\|_{Fr} = o(T^{-\epsilon})$ *and* $\|\tilde{\Xi}^-_p - \Xi^-_p\|_{Fr} = o(T^{-\epsilon})$ *(Ξ_+ and Ξ^-_p being non-singular) for some $\epsilon > 0$ and let $\tilde{\Xi}_+\hat{O}\hat{G}\tilde{\Xi}^-_p$ be obtained as the best (in Frobenius norm) rank n approximation of $\tilde{\Xi}_+\tilde{\beta}_r\tilde{\Xi}^-_p$. Further let* $O^\dagger := (O'(\Xi_+)^2O)^{-1}O'(\Xi_+)^2$ *assuming that* $\|(O'(\Xi_+)^2O)^{-1}\| < \infty$.

Then for T large enough $\hat{G}$ can a.s. be chosen such that $\hat{G}'S_p = I_n$. Further

$$\hat{G}' - G' = O^\dagger(\tilde{\beta}_r - b_r)(I_p - S_p G') + o(a_T). \tag{9}$$

PROOF: Since $\|\tilde{\beta}_r - b_r\|_{Fr} = o(a_T)$ it follows from the boundedness assumption on $O^\dagger$ that $O^\dagger\tilde{\beta}_r S_p \to O^\dagger \beta_r S_p = I_n$. Since $\hat{O}\hat{G}'$ is a best approximation to b_r based on $\tilde{\beta}_r$ in a weighted least squares sense it follows that

$$\|\hat{O}\hat{G}' - b_r\|_{Fr} \le \|\hat{O}\hat{G}' - \tilde{\beta}_r\|_{Fr} + \|\tilde{\beta}_r - b_r\|_{Fr} = O(\|\tilde{\beta}_r - b_r\|_{Fr}) = o(1).$$

It follows that $O^\dagger\hat{O}\hat{G}'S_p \to O^\dagger b_r S_p = I_n$. Therefore $\hat{G}$ subject to the restriction $\hat{G}'S_p = I_n$ is well defined a.s. for T large enough. It follows that $\|(\hat{O}\hat{G}' - b_r)S_p\|_{Fr} = \|\hat{O} - O\|_{Fr} = o(1)$. Letting $\hat{O}^\dagger := (\hat{O}'(\tilde{\Xi_+})^2\hat{O})^{-1}\hat{O}'(\tilde{\Xi_+})^2$ one obtains

$$\hat{G}' - G' = \hat{O}^\dagger\tilde{\beta}_r - O^\dagger b_r = (\hat{O}^\dagger - O^\dagger)\tilde{\beta}_r + O^\dagger(\tilde{\beta}_r - b_r).$$

Then $\|\hat{O} - O\| = o(1)$ and $\|\tilde{\Xi_+} - \Xi_+\| = o(T^{-\epsilon})$ together with the bounds on the norms $\|(O'(\Xi_+)^2O)^{-1}\|, \|\Xi_+\|$ and $\|O\|_{Fr}$ imply $\|O^\dagger - \hat{O}^\dagger\|_{Fr} = o(1)$. Using the fact that $(\hat{G}' - G')S_p = 0$ and $b_r(I - S_pG') = 0$ one obtains

$$\hat{G}' - G' = O^\dagger(\tilde{\beta}_r - b_r)(I - S_pG') + (\hat{O}^\dagger - O^\dagger)(\tilde{\beta}_r - b_r)(I - S_pG')$$

which proves the lemma. □

Lemma A.4 *Let $A_T = A_T' = A_0 + \delta^A, A_0 = A_0'$ and $B_T = B_0 + \delta^B$ be two sequences of matrices $A_T \in \mathbb{R}^{a\times a}$ and $B_T \in \mathbb{R}^{a\times b}$. A_0 and B_0 are possibly random matrices. Assume that all matrices are partitioned as*

$$\begin{aligned} A_T &= \begin{bmatrix} A_{T,11} & A_{T,12} \\ A_{T,21} & A_{T,22} \end{bmatrix} = \begin{bmatrix} A_{0,11} + \delta^A_{11} & \delta^A_{12} \\ \delta^A_{21} & A_{0,22} + \delta^A_{22} \end{bmatrix} = \begin{bmatrix} A_{0,11} + O_P(a_T^2) & O_P(a_T) \\ O_P(a_T) & A_{0,22} + O_P(a_T) \end{bmatrix}, \\ B_T &= \begin{bmatrix} B_{T,11} & B_{T,12} \\ B_{T,21} & B_{T,22} \end{bmatrix} = \begin{bmatrix} B_{0,11} + \delta^B_{11} & \delta^B_{12} \\ \delta^B_{21} & B_{0,22} + \delta^B_{22} \end{bmatrix} = \begin{bmatrix} B_{0,11} + O_P(b_T^2) & O_P(b_T) \\ O_P(b_T) & B_{0,22} + O_P(b_T) \end{bmatrix}, \end{aligned}$$

such that $A_{T,11} \in \mathbb{R}^{c\times c}, B_{T,11} \in \mathbb{R}^{c\times c}$ and all other matrices have the corresponding dimensions. The subscripts for all matrices indicate the corresponding blocks. Assume that $A_{0,11}^{-1} = O_P(1), A_{0,22}^{-1} = O_P(1)$. Finally let $J_T := B_T'(A_T)^{-1}B_T - B_0'(A_0)^{-1}B_0$.

Then if a_T and b_T are such that $a_T \to 0, b_T \to 0$ we have

$$
\begin{aligned}
J_{T,11} &= (\delta_{11}^B)' A_{0,11}^{-1} B_{0,11} + B_{0,11}' A_{0,11}^{-1} \delta_{11}^B - B_{0,11}' A_{0,11}^{-1} \delta_{11}^A A_{0,11}^{-1} B_{0,11} \\
&\quad + \left[(\delta_{21}^B)' - B_{0,11}' A_{0,11}^{-1} \delta_{12}^A \right] A_{0,22}^{-1} \left[\delta_{21}^B - \delta_{21}^A A_{0,11}^{-1} B_{0,11} \right] + o_P(a_T^2 + b_T^2) \\
J_{T,21} &= (\delta_{12}^B)' A_{0,11}^{-1} B_{0,11} + \left[B_{0,22} + \delta_{22}^B \right]' A_{0,22}^{-1} \left[\delta_{21}^B - \delta_{21}^A A_{0,11}^{-1} B_{0,11} \right] \\
&\quad - B_{0,22}' A_{0,22}^{-1} \delta_{22}^A A_{0,22}^{-1} \left[\delta_{21}^B - \delta_{21}^A A_{0,11}^{-1} B_{0,11} \right] + o_P(a_T^2 + b_T^2), \\
J_{T,22} &= (\delta_{22}^B - \delta_{21}^A A_{0,11}^{-1} \delta_{12}^B)' \left(A_{0,22}^{-1} - A_{0,22}^{-1} (\delta_{22}^A - \delta_{21}^A A_{0,11}^{-1} \delta_{12}^A - \delta_{22}^A A_{0,22}^{-1} \delta_{22}^A) A_{0,22}^{-1} \right) B_{0,22} \\
&\quad + B_{0,22}' \left(A_{0,22}^{-1} - A_{0,22}^{-1} (\delta_{22}^A - \delta_{21}^A A_{0,11}^{-1} \delta_{12}^A - \delta_{22}^A A_{0,22}^{-1} \delta_{22}^A) A_{0,22}^{-1} \right) (\delta_{22}^B - \delta_{21}^A A_{0,11}^{-1} \delta_{12}^B) \\
&\quad - B_{0,22}' \left(A_{0,22}^{-1} (\delta_{22}^A - \delta_{21}^A A_{0,11}^{-1} \delta_{12}^A - \delta_{22}^A A_{0,22}^{-1} \delta_{22}^A) A_{0,22}^{-1} \right) B_{0,22} \\
&\quad + (\delta_{22}^B - \delta_{21}^A A_{0,11}^{-1} \delta_{12}^B)' A_{0,22}^{-1} (\delta_{22}^B - \delta_{21}^A A_{0,11}^{-1} \delta_{12}^B) + (\delta_{12}^B)' A_{0,11}^{-1} \delta_{12}^B + o_P(a_T^2 + b_T^2).
\end{aligned}
$$

Therefore $J_{T,11} = O_P(a_T^2 + b_T^2)$ and $J_{T,12} = O_P(a_T + b_T), J_{T,22} = O_P(a_T + b_T)$. All evaluations hold if all in probability statements are exchanged by almost sure convergence.

PROOF: The proof follows from straightforward algebraic manipulations using the block matrix inversion

$$
A_T^{-1} = \left[\begin{array}{cc} A_{T,11}^{-1} & 0 \\ 0 & 0 \end{array} \right] + \left[\begin{array}{c} -A_{T,11}^{-1} A_{T,12} \\ I \end{array} \right] \left(A_{T,22} - A_{T,21} A_{T,11}^{-1} A_{T,12} \right)^{-1} \left[\begin{array}{c} -A_{T,21} A_{T,11}^{-1}, I \end{array} \right] \tag{10}
$$

noting that

$$
A_{T,11}^{-1} = (A_{0,11} + \delta_{11}^A)^{-1} = A_{0,11}^{-1} - A_{0,11}^{-1} \delta_{11}^A A_{0,11}^{-1} + o_P(a_T^2)
$$

since $a_T \to 0$ and $A_{0,11}^{-1} = O_P(1)$ by assumption. Similarly

$$
\left(A_{T,22} - A_{T,21} A_{T,11}^{-1} A_{T,12} \right)^{-1} = A_{0,22}^{-1} - A_{0,22}^{-1} \left(\delta_{22}^A - \delta_{21}^A A_{0,11}^{-1} \delta_{12}^A - \delta_{22}^A A_{0,22}^{-1} \delta_{22}^A \right) A_{0,22}^{-1} + o_P(a_T^2)
$$

follows. The remaining calculations are tedious but straightforward and hence omitted. □

Lemma A.5 *Define the two generalized eigenvalue problems:*

$$
(a) \quad \bar{Q}\bar{G} = \bar{M}\bar{G}\bar{R}^2, \quad (b) \quad \bar{\Phi}\bar{\Gamma} = \bar{\Psi}\bar{\Gamma}\bar{\Theta}^2.
$$

where $\bar{G} \in \mathbb{R}^{m_r \times n}, \bar{\Gamma} \in \mathbb{R}^{m_r \times n}, \bar{R}^2 \in \mathbb{R}^{n \times n}, \bar{\Theta} \in \mathbb{R}^{n \times n}$. Further $\bar{\Psi}$ and $\bar{\Theta}$ are assumed to be nonsingular a.s. and $\bar{\Theta}$ is diagonal.

(I) If $J := \bar{Q} - \bar{\Phi} = O(a_T)$ and $\delta_{zz} := \bar{M} - \bar{\Psi} = O(b_T)$ (where $a_T \to 0, b_T \to 0$ for $T \to \infty$) then there exists matrices $\bar{G}$ and $\bar{R}$ solving the eigenvalue problem (a) and matrices $\bar{\Gamma}$ and $\bar{\Theta}$

solving (b) such that $\bar{\Gamma}' S_p = I_n$ (where S_p denotes a selector matrix, i.e. a matrix consisting of columns of the identity matrix), $\bar{G} - \bar{\Gamma} = O(a_T + b_T), \bar{R} - \bar{\Theta} = O(a_T + b_T)$.

(II) Further let $\delta G := \bar{G} - \bar{\Gamma}$. Then the following two equations hold ($\bar{\Gamma}^\dagger := (\bar{\Gamma}'\bar{\Psi}\bar{\Gamma})^{-1}\bar{\Gamma}'$):

$$\begin{aligned} \bar{\Phi}\delta G - \bar{\Psi}\delta G \bar{R}^2 &= \delta_{zz}\bar{G}\bar{R}^2 + \bar{\Psi}\bar{\Gamma}(\bar{R}^2 - \bar{\Theta}^2) - J\bar{G}, & (11) \\ (I_m - \bar{\Psi}\bar{\Gamma}\bar{\Gamma}^\dagger)\bar{\Psi}\delta G\bar{R}^2 &= (I_m - \bar{\Psi}\bar{\Gamma}\bar{\Gamma}^\dagger)\left[J\bar{G} - \delta_{zz}\bar{G}\bar{R}^2 + \bar{\Phi}\delta G\right] & (12) \end{aligned}$$

(III) By transforming $\check{G} = \bar{G}(S_p'\bar{G})^{-1}$ it follows that $\check{G}$ solves the generalized eigenvalue problem (a) with matrix $\check{R}^2 = (S_p'\bar{G})\bar{R}^2(S_p'\bar{G})^{-1}$. Then $\check{G} - \bar{\Gamma} = O(a_T + b_T), \check{R} - \bar{\Theta} = O(a_T + b_T)$. Here $\check{R}$ is not necessarily block diagonal.

Proof: Solutions to the generalized eigenvalue problem are not identified. If all eigenvalues are distinct then fixing the sign of one nonzero entry in each column of $\bar{\Gamma}$ results in a unique solution (see e.g. Bauer et al., 1999, p. 1246, for a discussion). If there are repeated eigenvalues then more restrictions need to be introduced in order to achieve identification. It follows from operator theory (cf. e.g. Chatelin, 1983) that there exist normalizations such that the solution to the eigenvalue problem depends analytically on the matrix which is decomposed, i.e. such that $G - \Gamma = o(1)$ a.s. In these normalizations R^2 is not necessarily diagonal while still being block diagonal where the blocks correspond to the identical eigenvalues in Θ.

In particular let the sequence of matrices $\mathcal{M}_T \to \mathcal{M}_0$. Let $\tilde{\varphi}_i$ denote the matrix whose columns span the eigenspaces of $\mathcal{M}_T$ corresponding to the eigenvalues $\tilde{\lambda}_j \to \lambda_{0,i}, j = 1, \ldots, m_i$ where m_i denotes the multiplicity of the eigenvalue $\lambda_{0,i}$ of $\mathcal{M}_0$ with corresponding eigenspace spanned by the columns of the matrix $\varphi_{0,i}$. Here it is assumed that the normalization $\tilde{\varphi}_i'\varphi_{0,i} = I_{m_i} = \varphi_{0,i}'\varphi_{0,i}$ is chosen. Then it holds that

$$\tilde{\varphi}_i - \varphi_{0,i} = (\lambda_{0,i}I - \mathcal{M}_0)^\dagger(\mathcal{M}_T - \mathcal{M}_0)\varphi_{0,i} + O(\|\mathcal{M}_T - \mathcal{M}_0\|^2). \quad (13)$$

Here $X^\dagger$ denotes the Moore-Penrose pseudo-inverse. In particular let $\mathcal{M}_T = \bar{M}^{-1}\bar{Q}$ and $\mathcal{M}_0 = \bar{\Psi}^{-1}\bar{\Phi}$ and the columns of G and Γ equal $\tilde{\varphi}_i$ and $\varphi_{0,i}$ respectively. The condition of nonsingularity for $\bar{\Theta}$ ensures separation from the kernel of $\mathcal{M}_0$ and hence correct specification of the size of Γ. Then let $\bar{\Gamma} := \Gamma(S_p'\Gamma)^{-1}, \bar{G} := G(S_p'\Gamma)^{-1}$. Clearly $\bar{\Gamma}$ is a solution to the problem (b) fulfilling the assumption of the Lemma.

The assumptions imply that $\mathcal{M}_T - \mathcal{M}_0 = O(a_T + b_T)$ showing $G - \Gamma = O(a_T + b_T)$ and consequently $\bar{G} - \bar{\Gamma} = O(a_T + b_T)$. The order of convergence for $\bar{R} - \bar{\Theta}$ then follows from the fact that all other terms in (a) and (b) differ only by this order.

(II) Equation (11) follows from simple algebraic manipulations using the definitional equations (a) and (b) and $\bar{Q} = \bar{\Phi} + J, \bar{M} = \bar{\Psi} + \delta_{zz}, \bar{G} = \bar{\Gamma} + \delta G$. Premultiplying (11) with $\bar{\Gamma}^{\dagger}$ a rearranging of terms leads to

$$\bar{R}^2 - \bar{\Theta}^2 = \bar{\Gamma}^{\dagger}[\bar{\Phi}\delta G - \bar{\Psi}\delta G\bar{R}^2 - \delta_{zz}\bar{G}\bar{R}^2 + J\bar{G}]. \tag{14}$$

Inserting this into (11) shows that

$$(I_m - \bar{\Psi}\bar{\Gamma}\bar{\Gamma}^{\dagger})\left[\bar{\Phi}\delta G - \bar{\Psi}\delta G\bar{R}^2 - \delta_{zz}\bar{G}\bar{R}^2 + J\bar{G}\right] = 0.$$

This shows equation (12).

(III) Follows immediately from (I). □

A.2 Proof of Theorem 3.1

(I) Note that $\mathrm{diag}(\Delta(L)I_{c_r}, I_{m_r-c_r})H_r' z_t^r = c_v(L)\varepsilon_t$. Consequently the dimension of the cointegrating space of $(z_t^r)_{t\in\mathbb{N}}$ is equal to $m_r - c_r$. The claim on the dimension of the cointegrating space for $(y_t - b_u z_t^u)_{t\in\mathbb{Z}}$ also follows immediately from this representation.

(II) Note that $\tilde{y}_t$ denotes a transformation of $y_t - b_u z_t^u = b_r z_t^r + \Lambda\varepsilon_t = b_r H_r H_r' z_t^r + \Lambda\varepsilon_t$ which equals the estimation equation with the effects of z_t^u removed. Here we use that H_r was defined to be orthogonal. Next it is proved that matrices $\mathcal{T}_y$ and $\mathcal{T}_{z,r}$ transforming the equation into the required form exist.

Let $\mathcal{T}_y \in \mathbb{R}^{s\times s}$ and nonsingular $C \in \mathbb{R}^{c_r\times c_r}$ be chosen such that

$$\mathcal{T}_y b_r H_{r,\|} C = \begin{bmatrix} I_{c_y} & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{s\times c_r}.$$

It is easy to see that such choices always exist since c_y denotes the rank of $b_r H_{r,\|}$. Then in $\tilde{y}_t := \mathcal{T}_y(y_t - b_u z_t^u) = \mathcal{T}_y(b_r H_r H_r' z_t^r + \Lambda\varepsilon_t)$ the first c_y coordinates are integrated, the remaining being stationary. Choosing $\bar{\mathcal{T}}_{z,r} = \mathrm{diag}(C^{-1}, I_{m_r-c_r})H_r'$ we obtain that the first c_r components of $\bar{\mathcal{T}}_{z,r} z_t^r$ are integrated, the remaining ones being stationary. Using the above equation we obtain

$$\mathcal{T}_y b_r \bar{\mathcal{T}}_{z,r}^{-1} = \begin{bmatrix} I_{c_y} & 0 & \tilde{b}_{r,13} \\ 0 & 0 & \tilde{b}_{r,23} \end{bmatrix}.$$

Then the choice

$$\mathcal{T}_{z,r} = \begin{bmatrix} I & 0 & \tilde{b}_{r,13} \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} \bar{\mathcal{T}}_{z,r}$$

leads to the required representation. The remaining claims are straightforward to derive. Details are omitted.

A.3 Proof of Theorem 3.2

A.3.1 Consistency

Note that all estimators can be obtained in a two step procedure by first concentrating out z_t^u and afterwards maximizing the quasi likelihood with respect to β_r (Frisch-Waugh-Lovell equations). For fixed estimate $\hat{\beta}_r$ the least squares estimator for b_u is given by $\hat{\beta}_u = \langle y_t - \hat{\beta}_r z_t^r, z_t^u \rangle \langle z_t^u, z_t^u \rangle^{-1}$. Therefore

$$\hat{\beta}_u - b_u = \langle b_r z_t^r + b_u z_t^u + \Lambda \varepsilon_t - \hat{\beta}_r z_t^r - b_u z_t^u, z_t^u \rangle \langle z_t^u, z_t^u \rangle^{-1} = \langle \Lambda \varepsilon_t, z_t^u \rangle \langle z_t^u, z_t^u \rangle^{-1} + (b_r - \hat{\beta}_r) \langle z_t^r, z_t^u \rangle \langle z_t^u, z_t^u \rangle^{-1}. \tag{15}$$

This formula applies for the restricted and the unrestricted estimator. For the first term note that

$$\langle \varepsilon_t, z_t^u \rangle \langle z_t^u, z_t^u \rangle^{-1} = \langle \varepsilon_t, \mathcal{T}_{z,u} z_t^u \rangle \langle \mathcal{T}_{z,u} z_t^u, \mathcal{T}_{z,u} z_t^u \rangle^{-1} \mathcal{T}_{z,u} = \langle \varepsilon_t, \tilde{z}_t^u \rangle \langle \tilde{z}_t^u, \tilde{z}_t^u \rangle^{-1} \mathcal{T}_{z,u}.$$

Now (Bauer, 2009) implies that

$$\langle \varepsilon_t, \tilde{z}_t^u \rangle \langle \tilde{z}_t^u, \tilde{z}_t^u \rangle^{-1} = [O(P_T), O(Q_T)]$$

where $Q_T = \sqrt{\log\log T / T}$ and $P_T = \sqrt{\log T \log\log T / T^2}$.

In order to simplify the notation we use the symbols $\tilde{y}_t^\pi := \mathcal{T}_y (y_t - \langle y_t, z_t^u \rangle \langle z_t^u, z_t^u \rangle^{-1} z_t^u)$ and $\tilde{z}_t^\pi := \mathcal{T}_{z,r} (z_t^r - \langle z_t^r, z_t^u \rangle \langle z_t^u, z_t^u \rangle^{-1} z_t^u)$ throughout the proof. Here the superscript ^{r} corresponding to z_t^r will be omitted for notational simplicity. The corresponding symbols $\tilde{y}_{t,i}^\Pi$ and $\tilde{z}_{t,i}^\Pi$ denote the corresponding limit (a.s.) for $T \to \infty$ (where the symbols are only used if the limit exists). In general the residuals of the regression of any variable onto $z_t^u, t = 1, \dots, T$ will be denoted using the superscript $^\pi$ and $^\Pi$ will denote the corresponding limit (where it exists).

Using the same result and the Frisch-Waugh-Lovell equations for the first term and the orders of convergence stated in Lemma A.1

$$\begin{aligned}
\langle \tilde{\varepsilon}_t, \tilde{z}_t^\pi \rangle \langle \tilde{z}_t^\pi, \tilde{z}_t^\pi \rangle^{-1} &= [O(P_T), O(Q_T)], \\
\langle \tilde{z}_t, \tilde{z}_t^u \rangle \langle \tilde{z}_t^u, \tilde{z}_t^u \rangle^{-1} &= \langle \tilde{z}_t, \tilde{z}_t^u \rangle \mathrm{diag}(T^{-1} I, I) (\langle \tilde{z}_t^u, \tilde{z}_t^u \rangle \mathrm{diag}(T^{-1} I, I))^{-1} \\
&= \begin{bmatrix} O((\log\log T)^2) & O(\log T (\log\log T)^2) \\ O(1) & O(1) \end{bmatrix}.
\end{aligned}$$

Therefore it is sufficient to show that $(b_r - \hat{\beta}_r)$ converges to zero. To this end a transformed problem such that all transformed matrices converge to nonrandom matrices is analyzed first.

In this setting it will be possible to provide a.s. bounds for convergence rates. Afterwards the solution to the transformed problem is related to the solutions of the original problem.

Thus consider the transformed problem using the transformation matrices $\check{D}_y = \mathrm{diag}(\langle \tilde{z}_{t,1}, \tilde{z}_{t,1}\rangle^{-1/2}, I)$ and $\check{D}_r = \mathrm{diag}(\langle \tilde{z}_{t,1}, \tilde{z}_{t,1}\rangle^{-1/2}, I)$ respectively to transform the input and output of the estimated regression according to $\check{y}_t = \check{D}_y \tilde{y}_t, \check{z}_t = \check{D}_r \tilde{z}_t$. The transformed estimator $\check{\beta}_r := \check{D}_y \tilde{\beta}_{OLS,r} \check{D}_z^{-1}$ converges to $\tilde{b}_r = \mathcal{T}_y b_r \mathcal{T}_z^{-1} = OG'$ where the last equation defines O and G. Adapting the weighting $\check{\Xi}_+ := \tilde{\Xi}_+ \check{D}_y^{-1}$ we obtain $\Xi_+ = \mathrm{diag}(I_{c_y}, (\mathbb{E}\tilde{y}_{t,2}\tilde{y}_{t,2}')^{-1/2})$ and $\check{\Xi}_+ - \Xi_+ = O((\log T)\log\log T/\sqrt{T})$ as needed in Lemma A.3. This is obtained using the Cholesky factor as the square root of a matrix which is a differentiable operation. With this new normalization we obtain[2]

$$
\begin{aligned}
\delta\check{\beta}_r &:= \check{\beta}_{OLS,r} - \tilde{b}_r = \check{D}_y(\tilde{\beta}_{OLS,r} - \tilde{b}_r)\check{D}_z^{-1} = \check{D}_y[O(P_T), O(Q_T)]\check{D}_z^{-1} \\
&= \begin{bmatrix} O(\log T(\log\log T)^3/T) & O((\log\log T)^{3/2}/T) \\ O((\log T)^{3/2}/\sqrt{T}) & O(Q_T) \end{bmatrix}, \\
\check{\beta}_{RRR,r} - \tilde{b}_r &= \check{O}\check{G}' - OG' = (\check{O} - O)\check{G}' + O(\check{G}' - G'), \\
\check{O} - O &= (\check{\beta}_{RRR,r} - \tilde{b}_r)S_p, \\
\check{G}' - G' &= O^\dagger(\check{\beta}_{OLS,r} - \tilde{b}_r)(I - S_pG') + (\check{O}^\dagger - O^\dagger)(\check{\beta}_{OLS,r} - \tilde{b}_r)(I - S_pG')
\end{aligned}
$$

where

$$
S_p = \begin{bmatrix} I & 0 \\ 0 & 0 \\ 0 & I \\ 0 & 0 \end{bmatrix}, G = \begin{bmatrix} I & 0 \\ 0 & 0 \\ 0 & \Gamma_{3,2} \end{bmatrix}, I - S_pG' = I - \begin{bmatrix} I & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Gamma_{32}' \\ 0 & 0 & 0 \end{bmatrix}.
$$

Here the first two block rows of S_p correspond to $\tilde{z}_{t,1}$ and $\tilde{z}_{t,2}$. The remaining two blocks correspond to $\tilde{z}_{t,3}$. Since $\check{\beta}_{RRR,r}$ is a best rank n approximation to $\check{\beta}_r$ (see the proof of Lemma A.3) the rate of convergence of $\check{\beta}_r$ implies that $\check{\beta}_{RRR,r} - \tilde{b}_r = O((\log T)^{3/2}/\sqrt{T})$. Consequently also $\check{O} - O = O((\log T)^{3/2}/\sqrt{T})$ and hence also $\check{O}^\dagger - O^\dagger = O((\log T)^{3/2}/\sqrt{T})$ where $\check{O}^\dagger := (\check{O}'\check{\Xi}_+^2\check{O})^{-1}\check{O}'\check{\Xi}_+^2$. This shows the convergence rates for the solutions to the transformed problem. It thus remains to connect the solution to the original problem to the solution of the transformed problem.

It is straightforward to see that the untransformed estimate $\tilde{G}' = (\check{O}^\dagger \check{D}_y \tilde{\beta}_{OLS,r} S_p)^{-1} \check{O}^\dagger \check{D}_y \tilde{\beta}_{OLS,r}$ such that $\tilde{G}' S_p = I_n$. According to the limits above

$$
\check{O}^\dagger \check{D}_y \tilde{\beta}_{OLS,r} \check{D}_z^{-1} = \check{O}^\dagger \tilde{b}_r + \check{O}^\dagger(\check{D}_y \tilde{\beta}_{OLS,r} \check{D}_z^{-1} - \tilde{b}_r)
$$

[2]Here and below we will not always use the tightest possible bounds but use powers of $\log(T)$ instead for readability. Improvements are possible but their practical merits must be doubted.

$$= \check{O}^\dagger O G' + \check{O}^\dagger \delta \check{\beta}_r$$

Now let $\check{D}_n = \text{diag}(\langle \tilde{z}_{t,1}, \tilde{z}_{t,1}\rangle^{-1/2}, I_{n-c_y})$ such that $\check{D}_z S_p = S_p \check{D}_n$. Then

$$\begin{aligned}
[\tilde{G}' - G'] &= \check{D}_n^{-1}(\check{O}^\dagger O + \check{O}^\dagger \delta\check{\beta}_r S_p)^{-1}(\check{O}^\dagger O G' + \check{O}^\dagger \delta\check{\beta}_r)\check{D}_z - G' \\
&= \check{D}_n^{-1}(\check{O}^\dagger O + \check{O}^\dagger \delta\check{\beta}_r S_p)^{-1}((\check{O}^\dagger O + \check{O}^\dagger \delta\check{\beta}_r S_p)G' - \check{O}^\dagger \delta\check{\beta}_r S_p G' + \check{O}^\dagger \delta\check{\beta}_r)\check{D}_z - G' \\
&= \check{D}_n^{-1}(\check{O}^\dagger O + \check{O}^\dagger \delta\check{\beta}_r S_p)^{-1}\check{O}^\dagger \delta\check{\beta}_r (I - S_p G')\check{D}_z \\
&= \check{D}_n^{-1}(\check{O}^\dagger O + \check{O}^\dagger \delta\check{\beta}_r S_p)^{-1}\check{O}^\dagger \delta\check{\beta}_r \check{D}_z (I - S_p G') \\
&= \check{D}_n^{-1} O^\dagger \delta\check{\beta}_r \check{D}_z (I - S_p G') + \check{D}_n^{-1} O((\log T)^3/T)\check{D}_z \\
&= O^\dagger \check{D}_y^{-1} \delta\check{\beta}_r \check{D}_z (I - S_p G') + \check{D}_n^{-1} O((\log T)^3/T)\check{D}_z \\
&= [O((\log T)^4/T, (\log T)^4/\sqrt{T})]
\end{aligned}$$

where we have used that all terms have been shown above to be of order $O((\log T)^{3/2}/\sqrt{T})$. Further (due to the usage of the SVD and the corresponding orthogonality relations)

$$\begin{aligned}
\tilde{O} - O &= \tilde{O}\tilde{G}'\langle \tilde{z}_t^\pi, \tilde{z}_t^\pi\rangle \tilde{G}^\dagger - O = \tilde{\beta}_{OLS,r}\langle \tilde{z}_t^\pi, \tilde{z}_t^\pi\rangle \tilde{G}^\dagger - O \\
&= \langle \tilde{y}_t, \tilde{z}_t^\pi\rangle \tilde{G}^\dagger - O = \langle \tilde{b}_r \tilde{z}_t + \tilde{\varepsilon}_t, \tilde{z}_t^\pi\rangle \tilde{G}^\dagger - O \\
&= O G'\langle \tilde{z}_t^\pi, \tilde{z}_t^\pi\rangle \tilde{G}^\dagger - O + \langle \tilde{\varepsilon}_t, \tilde{z}_t^\pi\rangle \tilde{G}^\dagger \\
&= O \tilde{G}'\langle \tilde{z}_t^\pi, \tilde{z}_t^\pi\rangle \tilde{G}^\dagger - O + O(G' - \tilde{G}')\langle \tilde{z}_t^\pi, \tilde{z}_t^\pi\rangle \tilde{G}^\dagger + \langle \tilde{\varepsilon}_t, \tilde{z}_t^\pi\rangle \tilde{G}^\dagger \\
&= O(G' - \tilde{G}')\langle \tilde{z}_t^\pi, \tilde{z}_t^\pi\rangle \tilde{G}^\dagger + \langle \tilde{\varepsilon}_t, \tilde{z}_t^\pi\rangle \tilde{G}^\dagger
\end{aligned}$$

where

$$\tilde{G}^\dagger := \tilde{G}(\tilde{G}'\langle \tilde{z}_t^\pi, \tilde{z}_t^\pi\rangle \tilde{G})^{-1} = \tilde{G}\check{D}_n(\check{D}_n \tilde{G}'\langle \tilde{z}_t^\pi, \tilde{z}_t^\pi\rangle \tilde{G}\check{D}_n)^{-1}\check{D}_n.$$

It then follows from the orders of convergence provided in Lemma A.1 that $\tilde{D}_z\langle \tilde{z}_t^\pi, \tilde{z}_t^\pi\rangle \tilde{G}^\dagger \tilde{D}_n^{-1} = O(\log T)$. Consequently

$$\begin{aligned}
[\tilde{G}' - G']\langle \tilde{z}_t^\pi, \tilde{z}_t^\pi\rangle \tilde{G}^\dagger &= [O((\log T)^4/T), O((\log T)^4/\sqrt{T}]\tilde{D}_z^{-1}\tilde{D}_z\langle \tilde{z}_t^\pi, \tilde{z}_t^\pi\rangle \tilde{G}^\dagger \\
&= [O((\log T)^6/T), O((\log T)^6/\sqrt{T})].
\end{aligned}$$

Furthermore

$$\begin{aligned}
\langle \tilde{\varepsilon}_t^\pi, \tilde{z}_t^\pi\rangle \tilde{G}^\dagger &= \langle \tilde{\varepsilon}_t^\pi, \tilde{z}_t^\pi\rangle \langle \tilde{z}_t^\pi, \tilde{z}_t^\pi\rangle^{-1}\tilde{D}_z^{-1}\tilde{D}_z\langle \tilde{z}_t^\pi, \tilde{z}_t^\pi\rangle \tilde{G}^\dagger \\
&= [O(P_T), O(Q_T)]\tilde{D}_z^{-1} O(\log T)\tilde{D}_n \\
&= [O((\log T)^3/T), O((\log T)^3)/\sqrt{T})].
\end{aligned}$$

Together these orders imply $\tilde{O} - O = [O((\log T)^6/T), O((\log T)^6/\sqrt{T})]$ and thus

$$\tilde{\beta}_{RRR,r} - \tilde{b}_r = (\tilde{O} - O)G' + O(\tilde{G}' - G') + (\tilde{O} - O)(\tilde{G}' - G') = [O((\log T)^6/T), O((\log T)^6/\sqrt{T})].$$

Consequently we obtain from transforming (15)

$$\tilde{\beta}_{RRR,u} - \tilde{b}_u = O((\log T)^6/\sqrt{T}).$$

This shows the convergence rates.

A.3.2 Asymptotic Normality

In order to derive the asymptotic distribution of the estimator $\hat{\beta}_{RRR}$ the proof extends the theory contained in Anderson (2002). Since the proof is rather lengthy, the main steps are documented using lemmas summing up the main intermediate results.

Note that the RRR estimator is obtained from the singular value decomposition (using the symmetric matrix square roots)

$$\langle \tilde{y}_t^\pi, \tilde{y}_t^\pi \rangle^{-1/2} \langle \tilde{y}_t^\pi, \tilde{z}_t^\pi \rangle \langle \tilde{z}_t^\pi, \tilde{z}_t^\pi \rangle^{-1/2} = \hat{U}\hat{R}\hat{V}'.$$

Then as in Anderson (2002) (1.10), p. 205, the reduced rank estimator can be obtained as

$$\mathcal{T}_y \hat{\beta}_{RRR,r} \mathcal{T}_z^{-1} = \langle \tilde{y}_t^\pi, \tilde{z}_t^\pi \rangle \tilde{G} (\tilde{G}' \langle \tilde{z}_t^\pi, \tilde{z}_t^\pi \rangle \tilde{G})^{-1} \tilde{G}'$$

where $\tilde{G} = \langle \tilde{z}_t^\pi, \tilde{z}_t^\pi \rangle^{-1/2} \hat{V}_n \mathcal{T}_G \in \mathbb{R}^{m \times n}$ satisfies the following equations

$$\langle \tilde{z}_t^\pi, \tilde{y}_t^\pi \rangle \langle \tilde{y}_t^\pi, \tilde{y}_t^\pi \rangle^{-1} \langle \tilde{y}_t^\pi, \tilde{z}_t^\pi \rangle \tilde{G} = \langle \tilde{z}_t^\pi, \tilde{z}_t^\pi \rangle \tilde{G} (\mathcal{T}_G^{-1} \hat{R}^2 \mathcal{T}_G)$$

where $\hat{R}^2 = \text{diag}(\hat{r}_1^2, \hat{r}_2^2, \ldots, \hat{r}_n^2)$ denotes the matrix containing the squares of the n largest estimated singular values as its diagonal entries. The function of the transformation matrix $\mathcal{T}_G$ will become clear from the following.

Introduce the following notation (where in D_z the subscript $_r$ is omitted for notational simplicity):

$$\begin{aligned}
\tilde{D}_z &:= D_z T^{1/2} = \text{diag}(T^{-1/2} I, I), \quad \tilde{D}_y = D_y T^{1/2} = \text{diag}(T^{-1/2} I, I), \\
\bar{G} &:= \tilde{D}_z^{-1} \tilde{G}, \\
\bar{Q} &:= \langle \tilde{D}_z \tilde{z}_t^\pi, \tilde{D}_y \tilde{y}_t^\pi \rangle \langle \tilde{D}_y \tilde{y}_t^\pi, \tilde{D}_y \tilde{y}_t^\pi \rangle^{-1} \langle \tilde{D}_y \tilde{y}_t^\pi, \tilde{D}_z \tilde{z}_t^\pi \rangle, \\
\bar{M} &:= \langle \tilde{D}_z \tilde{z}_t^\pi, \tilde{D}_z \tilde{z}_t^\pi \rangle,
\end{aligned}$$

$$\bar{\Phi} := \begin{bmatrix} T^{-1}\langle \tilde{z}^{\pi}_{t,1}, \tilde{z}^{\pi}_{t,1}\rangle & T^{-1}\langle \tilde{z}^{\pi}_{t,1}, \tilde{z}^{\pi}_{t,2}\rangle & 0 \\ T^{-1}\langle \tilde{z}^{\pi}_{t,2}, \tilde{z}^{\pi}_{t,1}\rangle & T^{-1}\langle \tilde{z}^{\pi}_{t,2}, \tilde{z}^{\pi}_{t,1}\rangle\langle \tilde{z}^{\pi}_{t,1}, \tilde{z}^{\pi}_{t,1}\rangle^{-1}\langle \tilde{z}^{\pi}_{t,1}, \tilde{z}^{\pi}_{t,2}\rangle & 0 \\ 0 & 0 & \langle \tilde{z}^{\pi}_{t,3}, \tilde{y}^{\pi}_{t,2}\rangle\langle \tilde{y}^{\pi}_{t,2}, \tilde{y}^{\pi}_{t,2}\rangle^{-1}\langle \tilde{y}^{\pi}_{t,2}, \tilde{z}^{\pi}_{t,3}\rangle \end{bmatrix},$$

$$\bar{\Psi} := \begin{bmatrix} T^{-1}\langle \tilde{z}^{\pi}_{t,1}, \tilde{z}^{\pi}_{t,1}\rangle & T^{-1}\langle \tilde{z}^{\pi}_{t,1}, \tilde{z}^{\pi}_{t,2}\rangle & 0 \\ T^{-1}\langle \tilde{z}^{\pi}_{t,2}, \tilde{z}^{\pi}_{t,1}\rangle & T^{-1}\langle \tilde{z}^{\pi}_{t,2}, \tilde{z}^{\pi}_{t,2}\rangle & 0 \\ 0 & 0 & \langle \tilde{z}^{\pi}_{t,3}, \tilde{z}^{\pi}_{t,3}\rangle \end{bmatrix}.$$

A summary of the (unfortunately heavy) notation used can be found in Appendix B. The main guideline of the notation is to use Latin letters for matrices in which the stationary and the nonstationary subproblems are not separated (i.e. the off-diagonal blocks potentially are nonzero) and Greek letters for matrices for the decoupled problems. A bar indicates estimates (appropriately normalized so that convergence holds). This leads to two generalized eigenvalue problems related to SVDs:

$$(a) \quad \bar{Q}\bar{G} = \bar{M}\bar{G}\bar{R}^2, \quad (b) \quad \bar{\Phi}\bar{\Gamma} = \bar{\Psi}\bar{\Gamma}\bar{\Theta}^2.$$

Hence $\bar{G}$ denotes the solution to the original problem (a), $\bar{\Gamma}$ the solution to problem (b) where stationary and nonstationary components are separated. Consequently the solutions to (b) have the form:

$$\bar{\Gamma} = \begin{bmatrix} I & 0 \\ 0 & 0 \\ 0 & \bar{\Gamma}_{3,2} \end{bmatrix} \to \Gamma = \begin{bmatrix} I & 0 \\ 0 & 0 \\ 0 & \Gamma_{3,2} \end{bmatrix} \tag{16}$$

where the corresponding SVD for the stationary subproblem of (b) and its limit can be written as

$$\begin{aligned} \langle \tilde{z}^{\pi}_{t,3}, \tilde{z}^{\pi}_{t,3}\rangle\bar{\Upsilon}\bar{\Theta}_2 &= \langle \tilde{z}^{\pi}_{t,3}, \tilde{y}^{\pi}_{t,2}\rangle\langle \tilde{y}^{\pi}_{t,2}, \tilde{y}^{\pi}_{t,2}\rangle^{-1}\langle \tilde{y}^{\pi}_{t,2}, \tilde{z}^{\pi}_{t,3}\rangle\bar{\Upsilon}, \\ \mathbb{E}\tilde{z}^{\Pi}_{t,3}(\tilde{z}^{\Pi}_{t,3})'\Upsilon\Theta_2 &= \mathbb{E}\tilde{z}^{\Pi}_{t,3}(\tilde{y}^{\Pi}_{t,2})'(\mathbb{E}\tilde{y}^{\Pi}_{t,2}(\tilde{y}^{\Pi}_{t,2})')^{-1}\mathbb{E}\tilde{y}^{\Pi}_{t,2}(\tilde{z}^{\Pi}_{t,3})'\Upsilon. \end{aligned}$$

Solutions to these equation are not unique. In light of Lemma A.3 the restrictions $\Gamma_{3,2}'S_{p,22} = I = \bar{\Gamma}_{3,2}'S_{p,22}$ will be imposed. Here $S_{p,22}$ is a suitable selector matrix, i.e. a matrix whose columns are columns of an identity matrix. W.r.o.g. it can be assumed that $S_{p,22}' = [I, 0]$ by using an appropriate transformation $\mathcal{T}_z$. Note that this implies that Θ_2 and $\bar{\Theta}_2$ are not necessarily diagonal. Let $S_p = [S_{p,1}, S_{p,2}]$ where $S_{p,1}' = [I, 0]$ and $S_{p,2}' = [0, S_{p,22}']$. Then $\Gamma' S_p = I = \bar{\Gamma}' S_p$ are sufficient restrictions to identify the solutions Γ and $\bar{\Gamma}$. Analogously $\bar{G}_{3,2}'S_{p,22} = I, \bar{G}_{1,1} = I, \bar{R} = \text{diag}(\bar{R}_1, \bar{R}_2)$ identify a solution (asymptotically, see Lemma A.3 and Lemma A.5). These solutions will be used in the following. Here $\bar{\Theta}_2$ and Θ_2 resp. denote the $(2,2)$ blocks of $\bar{\Theta} = \text{diag}(I, \bar{\Theta}_2)$ and $\Theta = \text{diag}(I, \Theta_2)$ respectively.

The relations between the various solutions to the generalized eigenvalue problem are collected in section B. Throughout the rest of the proof we will use the following notation for

blocks of matrices: For a matrix X partitioned into blocks we let $X_{i,j}$ denote the blocks of the matrix. If multiple blocks are included also the notation $'i:j'$ will be used indicating the matrix built of blocks with indices i up to (and including) j. In order to denote block rows or columns we use a semicolon for selecting the whole row or column. Hence e.g. $\bar{G}_{3,2}$ denotes the (3,2) block, $\bar{G}_{1,:}$ the first block row and $\bar{G}_{1:2,1}$ the first two blocks rows in the first block column of the matrix $\bar{G}$.

The next lemma establishes orders of convergence of the solutions to the generalized eigenvalue problems.

Lemma A.6 *Let the assumptions of Theorem 3.2 hold.*

(I) Partition the matrices $\bar{Q}, \bar{M}, \bar{\Phi}, \bar{\Psi}$ according to the partitioning of $\tilde{z}_t$ denoting the various blocks using subscripts. Then

$$
\begin{aligned}
\delta_{zz} &:= \bar{M} - \bar{\Psi} = \begin{bmatrix} 0 & 0 & O_P(T^{-1/2}) \\ 0 & 0 & O_P(T^{-1/2}) \\ O_P(T^{-1/2}) & O_P(T^{-1/2}) & 0 \end{bmatrix}, \\
\delta_{yz} &:= \begin{bmatrix} T^{-1}(\langle \tilde{y}_{t,1}^{\pi}, \tilde{z}_{t,1}^{\pi}\rangle - \langle \tilde{z}_{t,1}^{\pi}, \tilde{z}_{t,1}^{\pi}\rangle) & T^{-1}(\langle \tilde{y}_{t,1}^{\pi}, \tilde{z}_{t,2}^{\pi}\rangle - \langle \tilde{z}_{t,1}^{\pi}, \tilde{z}_{t,2}^{\pi}\rangle) & T^{-1/2}\langle \tilde{y}_{t,1}^{\pi}, \tilde{z}_{t,3}^{\pi}\rangle \\ T^{-1/2}\langle \tilde{y}_{t,2}^{\pi}, \tilde{z}_{t,1}^{\pi}\rangle & T^{-1/2}\langle \tilde{y}_{t,2}^{\pi}, \tilde{z}_{t,2}^{\pi}\rangle & 0 \end{bmatrix} \\
&= \begin{bmatrix} O_P(T^{-1}) & O_P(T^{-1}) & O_P(T^{-1/2}) \\ O_P(T^{-1/2}) & O_P(T^{-1/2}) & 0 \end{bmatrix}, \\
\delta_{yy} &:= \begin{bmatrix} T^{-1}(\langle \tilde{y}_{t,1}^{\pi}, \tilde{y}_{t,1}^{\pi}\rangle - \langle \tilde{z}_{t,1}^{\pi}, \tilde{z}_{t,1}^{\pi}\rangle) & T^{-1/2}\langle \tilde{y}_{t,1}^{\pi}, \tilde{y}_{t,2}^{\pi}\rangle \\ T^{-1/2}\langle \tilde{y}_{t,2}^{\pi}, \tilde{y}_{t,1}^{\pi}\rangle & 0 \end{bmatrix} = \begin{bmatrix} O_P(T^{-1}) & O_P(T^{-1/2}) \\ O_P(T^{-1/2}) & 0 \end{bmatrix}.
\end{aligned}
$$

The terms $O_P(T^{-1})$ are $O((\log T)(\log\log T)^2/T)$ and the $O_P(T^{-1/2})$ terms are $O((\log T)(\log\log T)^2/T^{-1/2})$.

(II) Let $J := \bar{Q} - \bar{\Phi}$. To simplify notation define $Z_{ij} := T^{-1}\langle \tilde{z}_{t,i}^{\pi}, \tilde{z}_{t,j}^{\pi}\rangle, i,j = 1,2$. Then

$$
\begin{aligned}
J_{i,j} &= [\delta_{zy}^{i1} - Z_{i1}Z_{11}^{-1}\delta_{yy}^{11}]Z_{11}^{-1}Z_{1j} + Z_{i1}Z_{11}^{-1}\delta_{yz}^{1j} \\
&\quad + [\delta_{zy}^{i2} - Z_{i1}Z_{11}^{-1}\delta_{yy}^{12}](\mathbb{E}\tilde{y}_{t,2}^{\Pi}(\tilde{y}_{t,2}^{\Pi})')^{-1}[\delta_{yz}^{2j} - \delta_{yy}^{21}Z_{11}^{-1}Z_{1j}] + o_P(T^{-1}), \\
J_{3,i} &= \delta_{zy}^{31}Z_{11}^{-1}Z_{1i} + \langle \tilde{z}_{t,3}^{\pi}, \tilde{y}_{t,2}^{\pi}\rangle\langle \tilde{y}_{t,2}^{\pi}, \tilde{y}_{t,2}^{\pi}\rangle^{-1}[\delta_{yz}^{2i} - \delta_{yy}^{21}Z_{11}^{-1}Z_{1i}] + o_P(T^{-1}), \\
J_{3,3} &= [\langle \tilde{z}_{t,3}^{\pi}, \tilde{y}_{t,2}^{\pi}\rangle(\langle \tilde{y}_{t,2}^{\pi}, \tilde{y}_{t,2}^{\pi}\rangle)^{-1}\delta_{yy}^{21} - \delta_{zy}^{31}]\, Z_{11}^{-1}\, [\delta_{yy}^{12}(\langle \tilde{y}_{t,2}^{\pi}, \tilde{y}_{t,2}^{\pi}\rangle)^{-1}\langle \tilde{y}_{t,2}^{\pi}, \tilde{z}_{t,3}^{\pi}\rangle - \delta_{yz}^{13}] + o_P(T^{-1})
\end{aligned}
\tag{17}
$$

for $i = 1,2, j = 1,2$ where expressions for the remaining blocks of J follow from symmetry. Hence $J_{i,j} = O_P(T^{-1})$ and indeed $J_{i,j} = O((\log T)^3/T)$ for $i,j = 1,2$. Further $J_{3,i} = O_P(T^{-1/2})$ and indeed $J_{3,i} = O((\log T)^3/T^{-1/2})$ for $i = 1,2$. $J_{3,3} = O_P(T^{-1})$ and $J_{3,3} = O((\log T)^3/T)$ respectively.

(III) $\delta G := \bar{G} - \Gamma = O_P(T^{-1/2})$ and moreover $\delta G = O((\log T)^3/T^{1/2})$.

Proof: (I) The orders of convergence for the various entries of δ_{zz} and δ_{yz} follow from Lemma A.1. Details are omitted.

(II) Set $B_T := \tilde{D}_y \langle \tilde{y}_t^\pi, \tilde{z}_t^\pi \rangle \tilde{D}_z$ and $A_T := \tilde{D}_y \langle \tilde{y}_t^\pi, \tilde{y}_t^\pi \rangle \tilde{D}_y$ in Lemma A.4 where the partitioning refers to nonstationary and stationary components in the various matrices. Note that for the in probability part $a_T = b_T = T^{-1/2}$ and for the almost sure convergence $a_T = b_T = \log T (\log\log T)^2 / T^{-1/2}$ fulfill the assumptions of Lemma A.4. Also note that $\delta_{22}^A = 0, \delta_{22}^B = 0$ simplifying the expression for $J_{3,3}$. Then Lemma A.4 proves this part of the lemma. The orders of convergence for the various entries of J follow from the equations given and the orders of convergence provided in Lemma A.1. Here also the uniform bound on the two and infinity norm of $\langle \tilde{z}_{t,2}^u, \tilde{z}_{t,2}^u \rangle^{-1}, \langle \tilde{y}_{t,2}^\pi, \tilde{y}_{t,2}^\pi \rangle^{-1}$ which are implied by Assumption P are used.

(III) follows from (I) and (II) in combination with Lemma A.5. □

The next lemma (proven in section A.4) gathers more detailed results on the asymptotic properties of the entries of δG:

Lemma A.7 *Let the assumptions of Theorem 3.1 hold. Then we obtain*

$$\begin{aligned}
\delta G_{1,1} &= 0, \\
\delta G_{1,2} &= -Z_{11}^{-1} Z_{12} \delta G_{2,2} + Z_{11}^{-1} [\delta_{zz}^{13} \bar{\Gamma}_{3,2} \bar{\Theta}_2^2 - J_{1,3} \bar{\Gamma}_{3,2}] (I - \bar{\Theta}_2^2)^{-1} + o(T^{-1/2}), \\
\delta G_{2,1} &= (T^{-1} \langle \tilde{z}_{t,2.1}^\pi, \tilde{z}_{t,2.1}^\pi \rangle)^{-1} \left[J_{2.1,1} + (J_{2.1,3} - \delta_{zz}^{2.1,3}) \delta G_{3,1} \right] + o(T^{-1}) = O((\log T)^7 / T), \\
\delta G_{2,2} &= (T^{-1} \langle \tilde{z}_{t,2.1}^\pi, \tilde{z}_{t,2.1}^\pi \rangle)^{-1} \left[J_{2.1,3} \bar{\Gamma}_{3,2} \bar{\Theta}_2^{-2} - \delta_{zz}^{2.1,3} \bar{\Gamma}_{3,2} \right] + o(T^{-1/2}), \\
\delta G_{3,1} &= \bar{S} [J_{3,1} - \delta_{zz}^{31}] + o(T^{-1}), \\
\bar{P}_{3,3} \delta G_{3,1} &= (I - \langle \tilde{z}_{t,3}^\pi, \tilde{z}_{t,3}^\pi \rangle \bar{\Gamma}_{3,2} \bar{\Gamma}_{3,2}^\dagger) [J_{3,1} - \delta_{zz}^{31}] + o(T^{-1}) = O((\log T)^6 / T), \\
\delta G_{3,2} &= o(T^{-1/2})
\end{aligned}$$

where $\tilde{z}_{t,2.1}^\pi = \tilde{z}_{t,2}^\pi - Z_{21} Z_{11}^{-1} \tilde{z}_{t,1}^\pi, \delta_{zz}^{2.1,i} = \delta_{zz}^{2i} - Z_{21} Z_{11}^{-1} \delta_{zz}^{1i}, J_{2.1,1} = J_{1,1} - Z_{21} Z_{11}^{-1} J_{2,1}$,

$$\bar{S} = (Z_{33} - \langle \tilde{z}_{t,3}^\pi, \tilde{y}_{t,2}^\pi \rangle \langle \tilde{y}_{t,2}^\pi, \tilde{y}_{t,2}^\pi \rangle^{-1} \langle \tilde{y}_{t,2}^\pi, \tilde{z}_{t,3}^\pi \rangle)^{-1}$$

using $Z_{33} = \langle \tilde{z}_{t,3}^\pi, \tilde{z}_{t,3}^\pi \rangle$ *and*

$$\bar{P}_{33} = Z_{33} - Z_{33} \bar{\Gamma}_{3,2}^\dagger \bar{\Gamma}_{3,2}' Z_{33}. \tag{18}$$

Next these approximations are linked to the estimate $\hat{\beta}_{RRR,r}$.

Lemma A.8 *Let the assumptions of Theorem 3.2 hold.*

(I) Then

$$\mathcal{T}_y (\hat{\beta}_{RRR,r} - b_r) \mathcal{T}_z^{-1} D_z^{-1} = \left[\sqrt{T} \langle \tilde{\varepsilon}_t, \tilde{z}_t^\pi \rangle \tilde{D}_z \right] \bar{\Gamma} \bar{\Gamma}^\dagger + \begin{bmatrix} TI & 0 \\ 0 & \sqrt{T} \tilde{O}_2 \end{bmatrix} \delta G' (I - \bar{M} \bar{G} \bar{G}^\dagger)$$

$$+\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \sqrt{T}[\tilde\beta_{2,3}-\bar\beta_{2,3}][I-\mathbb{E}\tilde z^{\Pi}_{t,3}(\tilde z^{\Pi}_{t,3})'\Gamma_{3,2}\Gamma^{\dagger}_{3,2}] \end{bmatrix} + o_P(1)$$

where

$$\bar\beta_{2,3} = \langle \tilde y^{\pi}_{t,2}, \tilde z^{\pi}_{t,3}\rangle \bar\Gamma_{3,2}\bar\Gamma^{\dagger}_{3,2} = \bar{\mathcal{O}}_2\bar\Gamma'_{3,2} \to \tilde O_2 \Gamma'_{3,2} = \tilde b_{2,3}$$

denotes the solution to the subproblem of the problem (b) corresponding to the stationary components. $\Gamma^{\dagger}_{3,2} := (\Gamma'_{3,2}\mathbb{E}\tilde z^{\pi}_{t,3}(\tilde z^{\pi}_{t,3})'\Gamma_{3,2})^{-1}\Gamma'_{3,2}$.

(II) Letting $\delta G_{:,1}$ and $\delta G_{:,2}$ denote the first and second block column of δG it holds that

$$T\delta G'_{:,1}(I-\bar M\bar G\bar G^{\dagger}) = \begin{bmatrix} -T\delta H' Z_{21}Z_{11}^{-1} & T\delta H' & T\delta G'_{3,1}\bar P \end{bmatrix} + o(1), \qquad (19)$$

$$\sqrt{T}\delta G'_{:,2}(I-\bar M\bar G\bar G^{\dagger}) = \sqrt{T}\delta G'_{2,2}\begin{bmatrix} -Z_{21}Z_{11}^{-1} & I & 0 \end{bmatrix} + o(1) \qquad (20)$$

where $\delta H = \delta G_{2,1} - \delta G_{2,2}(\Gamma^{\dagger}_{3,2})'\mathbb{E}\tilde z^{\Pi}_{t,3}(\tilde z^{\Pi}_{t,3})'\delta G_{3,1}$ The term $o_P(1)$ is also $O((\log T)^a)$.

Part (I) of the lemma splits the estimation error $\hat\beta_{RRR,r} - b_r$ (up to errors of higher order) into three terms: The first term asymptotically equals the estimation error in the situation that the row space of b_r is known and used in the estimation to obtain unrestricted least squares estimators. The second term accounts for the effects of the rank restriction in the nonstationary components of both the output (i.e. $\tilde y^{\pi}_{t,1}$) as well as the regressors (i.e. $\tilde z^{\pi}_{t,i}, i = 1,2$). Part (II) of the lemma provides a more detailed expression for this term. The last term corrects for the rank restriction in the stationary directions.

Up to now only for showing $\sqrt{T}\delta G'_{3,1}(I - \langle \tilde z^{\pi}_{t,3}, \tilde z^{\pi}_{t,3}\rangle\bar\Gamma_{3,2}\bar\Gamma^{\dagger}_{32}) \to 0$ the exact form of δ_{zz} and J are used. All other results up to now rely only on the order of convergence of these terms. The proof of Theorem 3.2 is completed with the last lemma of this section giving explicit expressions for the asymptotic distributions of the various parts of the expressions for $\mathcal{T}_y(\hat\beta_{RRR,r} - b_r)\mathcal{T}_{z,r}^{-1}D_{z,r}^{-1}$ given in Lemma A.8. The result then follows directly from (15).

Lemma A.9 *With $\tilde O_2^{\dagger} = (\tilde O_2'(\mathbb{E}\tilde y^{\Pi}_{t,2}(\tilde y^{\Pi}_{t,2})')^{-1}\tilde O_2)^{-1}\tilde O_2'(\mathbb{E}\tilde y^{\Pi}_{t,2}(\tilde y^{\Pi}_{t,2})')^{-1}$ we have*

$$\begin{aligned}
\langle \tilde\varepsilon_t, \tilde z^{\pi}_{t,1}\rangle Z_{11}^{-1} &\stackrel{d}{\to} f(\mathcal{T}_y\Lambda W, W^{\Pi}_{z,1}),\\
\sqrt{T}\langle \tilde\varepsilon_t, \tilde z^{\pi}_{t,3}\rangle\langle \tilde z^{\pi}_{t,3}, \tilde z^{\pi}_{t,3}\rangle^{-1} &\stackrel{d}{\to} Z_r \sim \mathcal{N}(0,V),\\
\sqrt{T}\delta G_{2,2} &= (T^{-1}\langle \tilde z^{\pi}_{t,2.1}, \tilde z^{\pi}_{t,2.1}\rangle)^{-1}\langle \tilde z^{\pi}_{t,2.1}, \tilde\varepsilon_{t,2}\rangle(\tilde O_2^{\dagger})' + o_P(1) \stackrel{d}{\to} M'_{r,2}(\tilde O_2^{\dagger})',\\
T\bar P'\delta G_{3,1} &= Z_{33}^{-1}P\sqrt{T}\langle \tilde z^{\pi}_{t,3}, \tilde\varepsilon_{t,1}\rangle + Z_{33}^{-1}\sqrt{T}\left(\bar P\langle \tilde z^{\pi}_{t,3}, \tilde y_{t,2}\rangle - P\mathbb{E}\tilde z^{\Pi}_{t,3}(\tilde y^{\Pi}_{t,2})'\right)\Xi' + o_P(1) \to \tilde R',\\
T\delta H &= (T^{-1}\langle \tilde z^{\pi}_{t,2.1}, \tilde z^{\pi}_{t,2.1}\rangle)^{-1}\left[\langle \tilde z^{\pi}_{t,2.1}, \tilde\varepsilon_{t,1}\rangle + \langle \tilde z^{\pi}_{t,2.1}, \tilde\varepsilon^{\pi}_{t,2}\rangle(I-\tilde O_2\tilde O_2^{\dagger})'\Xi'\right] + o_P(1) \stackrel{d}{\to} N',\\
\sqrt{T}[\bar\beta_{2,3} - \tilde\beta_{2,3}]P &= \tilde O_2\tilde O_2^{\dagger}\sqrt{T}\langle \tilde\varepsilon_{t,2}, \tilde z^{\Pi}_{t,3}\rangle(\mathbb{E}\tilde z^{\Pi}_{t,3}(\tilde z^{\Pi}_{t,3})')^{-1}P + o_P(1),
\end{aligned}$$

where $P = (I - \mathbb{E}\tilde{z}^{\Pi}_{t,3}(\tilde{z}^{\Pi}_{t,3})'\Gamma_{3,2}\Gamma^{\dagger}_{3,2})$. *Further* $Y_{i1} = \int W^{\Pi}_{z,i}(W^{\Pi}_{z,1})'$. *Here* $M_{r,2} = [0, I]M_r$ *with* $M_r := f(\mathcal{T}_y \Lambda W, W^{\Pi}_{z,2.1})$ *where* $W_{z,2.1} := W^{\Pi}_{z,1} - Y_{21}Y_{11}^{-1}W^{\Pi}_{z,1}$. $N = [[I, 0] + \Xi(I - \tilde{O}_2\tilde{O}_2^{\dagger})[0, I]]M_r$. $T\bar{P}'\delta G_{3,1}$ *and* $\sqrt{T}[\tilde{\beta}_{2,3} - \beta_2]P$ *converge in distribution to Gaussian random variables with mean zero.*

Combining Lemma A.8 and A.9 we obtain that $\mathcal{T}_y(\hat{\beta}_{RRR,r} - b_r)\mathcal{T}_{z,r}^{-1}D_z^{-1} \xrightarrow{d}$

$$\left[f(\mathcal{T}_y \Lambda W, W^{\Pi}_{z,1}), 0, Z_r Z_{33}\Gamma_{3,2}\Gamma^{\dagger}_{3,2}\right] + \left[\begin{array}{ccc} -NY_{21}Y_{11}^{-1} & N & \tilde{R} \\ -\tilde{O}_2\tilde{O}_2^{\dagger}M_{r,2}Y_{21}Y_{11}^{-1} & \tilde{O}_2\tilde{O}_2^{\dagger}M_{r,2} & \tilde{O}_2\tilde{O}_2^{\dagger}Z_{r,2}P \end{array}\right].$$

From this the claim follows using the block matrix inversion since $\mathcal{T}_y(\hat{\beta}_{OLS,r} - b_r)\mathcal{T}_z^{-1}D_z^{-1} \xrightarrow{d} [f(\mathcal{T}_y \Lambda W, W^{\Pi}_z), Z_r]$.

Using standard asymptotics for the term $\langle \varepsilon_t, z_t^u\rangle\langle z_t^u, z_t^u\rangle^{-1}$ we obtain the asymptotic distribution of $\tilde{\beta}_{RRR,u} - \tilde{\beta}_{OLS,u}$. stated in Theorem 3.2 from (15). Note in particular that $R := Z_{r,1}P - \tilde{R}$ where $Z_{r,1}$ denotes the limit of $\sqrt{T}\langle\tilde{\varepsilon}_{t,1}, \tilde{z}^{\Pi}_{t,3}\rangle Z_{33}^{-1}$. This concludes the proof of (I) of Theorem 3.2.

Following the arguments of the proof and using Lemma A.1 (IV) it follows that the usual changes occur if a constant (and a deterministic trend respectively) is included in the regression: The asymptotics in the stationary directions are unchanged. For the nonstationary directions the Brownian motions are replaced by their corresponding demeaned (and detrended respectively) versions. We omit details in this respect.

A.4 Proof of Lemma A.7

In the proof the following results are used: With $\Xi := -\mathbb{E}\tilde{\varepsilon}_{t,1}\tilde{y}_{t,2}'(\mathbb{E}\tilde{y}^{\Pi}_{t,2}(\tilde{y}^{\Pi}_{t,2})')^{-1}$, $\tilde{z}_{t,2.1} = \tilde{z}_{t,2} - Z_{21}Z_{11}^{-1}\tilde{z}_{t,1}$, $J_{2.1,1} = J_{2,1} - Z_{21}Z_{11}^{-1}J_{1,1}$, $J_{2.1,3} = J_{2,3} - Z_{21}Z_{11}^{-1}J_{1,3}$ and $\tilde{\varepsilon}_{t,1.2} = \tilde{\varepsilon}_{t,1} + \Xi\tilde{y}^{\Pi}_{t,2}$ it holds that

$$\begin{aligned}
\sqrt{T}[\delta^{21}_{yz} - \delta^{21}_{yy}]'\langle\tilde{y}^{\pi}_{t,2}, \tilde{y}^{\pi}_{t,2}\rangle^{-1} &= \Xi + o(1), \\
\sqrt{T}[J_{3,1} - \delta^{3,1}_{zz}] &= \mathbb{E}\tilde{z}^{\Pi}_{t,3}(\tilde{y}^{\Pi}_{t,2})'\Xi' + o_P(1), \\
TJ_{2.1,1} &= \langle\tilde{z}^{\pi}_{t,2.1}, \tilde{\varepsilon}_{t,1.2}\rangle + o_P(1), \\
\sqrt{T}J_{2.1,3} &= \langle\tilde{z}^{\pi}_{t,2.1}, \tilde{y}^{\pi}_{t,2}\rangle(\mathbb{E}\tilde{y}^{\Pi}_{t,2}(\tilde{y}^{\Pi}_{t,2})')^{-1}\tilde{O}_2\Gamma'_{3,2}Z_{33} + o_P(1).
\end{aligned}$$

where $Z_{33} := \mathbb{E}\tilde{z}^{\Pi}_{t,3}(\tilde{z}^{\Pi}_{t,3})'$.

The first claim follows from

$$\sqrt{T}[\delta^{21}_{yz} - \delta^{21}_{yy}] = \langle\tilde{y}^{\pi}_{t,2}, \tilde{z}^{\pi}_{t,1} - \tilde{y}^{\pi}_{t,1}\rangle = -\langle\tilde{y}^{\pi}_{t,2}, \tilde{\varepsilon}_{t,1}\rangle \to -\mathbb{E}\tilde{\varepsilon}_{t,2}\tilde{\varepsilon}'_{t,1}$$

where $\tilde{y}^{\pi}_{t,2} = \tilde{\varepsilon}_{t,2} + \tilde{b}_{2,3}\tilde{z}_{t,3} - \langle \tilde{\varepsilon}_{t,2} + \tilde{b}_{2,3}\tilde{z}_{t,3}, z^u_t \rangle \langle z^u_t, z^u_t \rangle^{-1} z^u_t$. Then the result follows straightforwardly since $\tilde{z}_{t,3}$ and $\tilde{\varepsilon}_{t,1}$ are stationary and uncorrelated by assumption. From this also the second and third claim follow immediately using the expressions for $J_{i,1}$ derived in Lemma A.7. The fourth claim also follows from these expressions noting that $\mathbb{E}\tilde{y}^{\Pi}_{t,2}(\tilde{z}^{\Pi}_{t,3})' = \tilde{O}_2 \Gamma'_{3,2} \mathbb{E}\tilde{z}^{\Pi}_{t,3}(\tilde{z}^{\Pi}_{t,3})'$.

Now with respect to the blocks of δG note that due to the chosen normalizations $\bar{\Gamma}_{1,1} = \bar{G}_{1,1} = I$ implying $\delta G_{1,1} = 0$.

Using the order of convergence for J, δ_{zz} (Lemma A.6 (I) and (II)) and δG (Lemma A.6 (III)) the (1,2) block of (11) implies

$$\delta G_{1,2} + Z_{11}^{-1} Z_{12} \delta G_{2,2} = Z_{11}^{-1}[\delta^{13}_{zz} \bar{G}_{3,2} \bar{R}^2_2 - J_{1,3}\bar{G}_{3,2}](I - \bar{R}^2_2)^{-1} + o(T^{-1}).$$

The expression given then follows from noting that $\delta^{13}_{zz} = O((\log T)^a / T^{1/2}), \bar{G}_{3,2} = \bar{\Gamma}_{3,2} + O((\log T)^3/T^{1/2}), J_{1,3} = O((\log T)^3/T^{1/2})$ and $\bar{R}^2_2 = \bar{\Theta}^2_2 + O((\log T)^3/T^{1/2})$.

The expressions for $\delta G_{2,1}$ and $\delta G_{2,2}$ follow from the second block row of equation (12) noting that

$$\begin{aligned}
(I_m - \bar{\Psi}\bar{\Gamma}^{\dagger}\bar{\Gamma}') &= \begin{bmatrix} 0 & 0 & 0 \\ -Z_{21}Z_{11}^{-1} & I & 0 \\ 0 & 0 & I - \bar{\Psi}_{3,3}\bar{\Gamma}_{3,2}(\bar{\Gamma}'_{3,2}\bar{\Psi}_{3,3}\bar{\Gamma}_{3,2})^{-1}\bar{\Gamma}'_{3,2} \end{bmatrix}, \\
(I_m - \bar{\Psi}\bar{\Gamma}\bar{\Gamma}^{\dagger})\bar{\Psi} &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & T^{-1}\langle \tilde{z}^{\pi}_{t,2.1}, \tilde{z}^{\pi}_{t,2.1} \rangle & 0 \\ 0 & 0 & \bar{P}_{33} \end{bmatrix}.
\end{aligned}$$

Also $\bar{R}^2_1 - I = O((\log T)^7/T)$ follows from the $(1,1)$ entry of (11). Then the $(3,1)$ entry of equation (11) implies that

$$\delta G_{3,1} = \bar{S}[J_{3,1} - \delta^{31}_{zz}] + o(T^{-1})$$

where $\bar{S}^{-1} \to S^{-1}$ as

$$\langle \tilde{z}^{\pi}_{t,3}, \tilde{z}^{\pi}_{t,3} \rangle - \langle \tilde{z}^{\pi}_{t,3}, \tilde{y}^{\pi}_{t,2} \rangle \langle \tilde{y}^{\pi}_{t,2}, \tilde{y}^{\pi}_{t,2} \rangle^{-1} \langle \tilde{y}^{\pi}_{t,2}, \tilde{z}^{\pi}_{t,3} \rangle \to \mathbb{E}\tilde{z}^{\Pi}_{t,3}(\tilde{z}^{\Pi}_{t,3})' - \mathbb{E}\tilde{z}^{\Pi}_{t,3}(\tilde{y}^{\Pi}_{t,2})'(\mathbb{E}\tilde{y}^{\Pi}_{t,2}(\tilde{y}^{\Pi}_{t,2})')^{-1}\mathbb{E}\tilde{y}^{\Pi}_{t,2}(\tilde{z}^{\Pi}_{t,3})' > 0$$

which is ensured by the assumed nonsingularity of the covariance of $[(\tilde{y}^{\Pi}_{t,2})', (\tilde{z}^{\Pi}_{t,3})']'$. Further the $(3,1)$ entry in equation (12) directly implies the expression for $\bar{P}_{3,3}\delta G_{3,1}$ using the orders of convergence derived above. Next

$$J_{3,1} - \delta^{31}_{zz} = \delta^{31}_{zy} - \delta^{31}_{zz} + \langle \tilde{z}^{\pi}_{t,3}, \tilde{y}^{\pi}_{t,2} \rangle \langle \tilde{y}^{\pi}_{t,2}, \tilde{y}^{\pi}_{t,2} \rangle^{-1}(\delta^{21}_{yz} - \delta^{21}_{yy}) = T^{-1/2}[\langle \tilde{z}^{\pi}_{t,3}, \tilde{\varepsilon}_{t,1} \rangle + \langle \tilde{z}^{\pi}_{t,3}, \tilde{y}^{\pi}_{t,2} \rangle (\Xi' + o(1))].$$

Note that $\langle \tilde z_{t,3}^\pi, \tilde\varepsilon_{t,1}\rangle = O(Q_T)$, $Z_{33}^{-1}\langle \tilde z_{t,3}^\pi, \tilde y_{t,2}^\pi\rangle = \Gamma_{3,2}\tilde O_2' + O(Q_T)$, $\bar P - P = O((\log T)^3/T^{1/2}))$ and thus $\sqrt{T}\bar P_{3,3}\delta G_{3,1} =$

$$= \sqrt{T}\bar P(J_{31} - \delta_{zz}^{31}) + o(T^{-1/2}) = \bar P\langle \tilde z_{t,3}^\pi, \tilde\varepsilon_{t,1}\rangle + \bar P\langle \tilde z_{t,3}^\pi, \tilde y_{t,2}^\pi\rangle(\Xi' + o(1)) + o(T^{-1/2}) = O((\log T)^6/T^{1/2}).$$

Here the last order follows from $P\mathbb{E}\tilde z_{t,3}^\Pi(\tilde z_{t,3}^\Pi)'\Gamma_{3,2} = 0$ as is easy to verify.

Since $\delta G = O((\log T)^3)/T^{1/2})$ also $\delta G_{1,2} = O((\log T)^3/T^{1/2})$ and $\delta G_{2,2} = O((\log T)^3/T^{1/2})$. For $\bar\Gamma_\perp$ such that $\bar\Gamma_\perp'\bar\Gamma_{3,2} = 0$ we have that $\bar\Gamma_\perp' Z_{33}^{-1}\bar P_{33} = \bar\Gamma_\perp'$. Then the (3,2) block of (12) shows that

$$[0, \bar\Gamma_\perp']\delta G_{:,2} = \bar\Gamma_\perp'\delta G_{3,2} = \bar\Gamma_\perp' Z_{33}^{-1}\bar P_{33}\delta G_{3,2} = [0, \bar\Gamma_\perp' Z_{33}^{-1}][J\bar G_{:,2}\bar R_2^{-2} - \delta_{zz}\bar G_{:,2}] = o(T^{-1/2})$$

since the (3,2) block entry of $J\bar G$ and $\delta_{zz}\bar G$ both are of order $o(T^{-1/2})$ as follows from the norm bounds given for the blocks of J and δ_{zz}.

Due to the chosen normalization $\bar\Gamma_{3,2}' S_{p,22} = \bar G_{3,2}' S_{p,22} = I$ and thus $S_{p,22}'\delta G_{3,2} = 0$. Since $[\bar\Gamma_\perp, S_{p,22}]$ is nonsingular (as is straightforward to see from $\Gamma_{3,2}' S_{p,22} = I$) we obtain $\delta G_{3,2} = o(T^{-1/2})$.

The order of convergence for $\delta G_{2,1}$ follows from the orders of convergence of J, δ_{zz} and δG as derived in Lemma A.6.

A.5 Proof of Lemma A.8

(I) Let $\tilde\beta_{RRR,r} = \mathcal{T}_y\hat\beta_{RRR,r}\mathcal{T}_{z,r}^{-1}$ and $\tilde b_r = \mathcal{T}_y b_r \mathcal{T}_{z,r}^{-1}$. Then

$$\begin{aligned}\tilde\beta_{RRR,r} &= \langle \tilde y_t^\pi, \tilde z_t^\pi\rangle \hat G\hat G^\dagger = \tilde b_r\langle \tilde z_t^\pi, \tilde z_t^\pi\rangle \tilde D_z\bar G\bar G^\dagger\tilde D_z' + \langle \tilde\varepsilon_t, \tilde z_t^\pi\rangle\tilde D_z\bar G\bar G^\dagger\tilde D_z \\ &\doteq \tilde b_r\langle \tilde z_t^\pi, \tilde z_t^\pi\rangle \tilde D_z\bar G\bar G^\dagger\tilde D_z' + \left[\sqrt{T}\langle \tilde\varepsilon_t, \tilde z_t^\pi\rangle\tilde D_z\right]\bar\Gamma\bar\Gamma^\dagger D_z\end{aligned}$$

where $\doteq$ stands for equality up to terms of order $o(T^{-1})$ in the first c_z columns and of order $o(T^{-1/2})$ in the remaining columns. This follows from $\langle \tilde\varepsilon_t, \tilde z_t^\pi\rangle\tilde D_z = O((\log T)^3)T^{-1/2}), \delta G = O((\log T)^3 T^{-1/2})$ and the definition of $D_z = \mathrm{diag}(T^{-1}I, T^{-1/2}I)$ showing that $\bar G$ can be replaced by $\bar\Gamma$ in the second term in the last equation with introduction of an error of the stated order. Now since $\tilde b_r$ is block diagonal

$$\tilde b_r\langle \tilde z_t^\pi, \tilde z_t^\pi\rangle\tilde D_z\bar G\bar G^\dagger\tilde D_z = \tilde D_y^{-1}\tilde b_r\tilde D_z\langle \tilde z_t^\pi, \tilde z_t^\pi\rangle\tilde D_z\bar G\bar G^\dagger\tilde D_z.$$

Next note that (using the index ':' to denote block columns or rows resp.) $\tilde b_{r,1,:} = [I, 0] = \bar\Gamma_{:,1}'$. Therefore we have recalling that $(\bar G^\dagger)' = \bar G(\bar G'\bar M\bar G)^{-1}$

$$\begin{aligned}[I, 0] &= \bar G_{:,1}'\bar M(\bar G^\dagger)' = (\bar\Gamma_{:,1})'\bar M(\bar G^\dagger)' + \delta G_{:,1}'\bar M(\bar G^\dagger)' \\ &= \tilde b_{r,1,:}\bar M(\bar G^\dagger)' + \delta G_{:,1}'\bar M(\bar G^\dagger)'.\end{aligned}$$

This leads to $\tilde{b}_{r,1,:}\bar{M}(\bar{G}^\dagger)' = [[I,0] - \delta G'_{:,1}\bar{M}(\bar{G}^\dagger)']$. Hence we obtain

$$\begin{aligned}\tilde{b}_{r,1,:}\langle \tilde{z}_t^\pi, \tilde{z}_t^\pi\rangle \tilde{D}_z\bar{G}\bar{G}^\dagger\tilde{D}_z &= T^{1/2}\left[[I,0] - \delta G'_{:,1}\bar{M}(\bar{G}^\dagger)'\right]\bar{G}'\tilde{D}_z \\ &= \tilde{b}_{r,1,:} + T\left[\delta G'_{:,1}\right]\left[I - \bar{M}\bar{G}\bar{G}^\dagger\right]D_z.\end{aligned}$$

Here $\bar{G}'_{:,1} = [I,0] + \delta G'_{:,1}$ is used in the last line.

With respect to the second block row it follows from $\tilde{b}_{r,2,:} = [0, \tilde{b}_{2,3}]$ that $\tilde{b}_{r,2,:} = \tilde{b}_{r,2,:}\tilde{D}_z$. Also we have from $\bar{\beta}_{2,3} = \bar{\mathcal{O}}_2\bar{\Gamma}'_{3,2}$ using $\bar{G}'_{:,2}\bar{M}\bar{G}\bar{G}^\dagger = \bar{G}'_{:,2}$ by the definition of $\bar{G}^\dagger$ that

$$\begin{aligned}\bar{\beta}_{2,3}\langle \tilde{z}_{t,3}^\pi, \tilde{z}_t^\pi\rangle \tilde{D}_z\bar{G}\bar{G}^\dagger\tilde{D}_z &= \bar{\mathcal{O}}_2\bar{\Gamma}'_{3,2}\langle \tilde{z}_{t,3}^\pi, \tilde{z}_t^\pi\rangle \tilde{D}_z\bar{G}\bar{G}^\dagger\tilde{D}_z = \bar{\mathcal{O}}_2\bar{\Gamma}'_{:,2}\bar{M}\bar{G}\bar{G}^\dagger\tilde{D}_z \\ &= \bar{\mathcal{O}}_2[\bar{G}'_{:,2}\tilde{D}_z - \delta G'_{:,2}\bar{M}\bar{G}\bar{G}^\dagger\tilde{D}_z] = \bar{\mathcal{O}}_2[\bar{\Gamma}'_{:,2} + \delta G'_{:,2}(I - \bar{M}\bar{G}\bar{G}^\dagger)\tilde{D}_z] \\ &= [0, \bar{\beta}_{2,3}] + \bar{\mathcal{O}}_2\delta G'_{:,2}(I - \bar{M}\bar{G}\bar{G}^\dagger)\tilde{D}_z.\end{aligned}$$

This implies

$$\begin{aligned}\tilde{b}_{r,2,:}\langle \tilde{z}_t^\pi, \tilde{z}_t^\pi\rangle \tilde{D}_z\bar{G}\bar{G}^\dagger\tilde{D}_z &= \tilde{b}_{r,2,:}\tilde{D}_z\langle \tilde{z}_t^\pi, \tilde{z}_t^\pi\rangle \tilde{D}_z\bar{G}\bar{G}^\dagger\tilde{D}_z \\ &= [\tilde{b}_{2,3} - \bar{\beta}_{2,3}]\langle \tilde{z}_{t,3}^\pi, \tilde{z}_t^\pi\rangle \tilde{D}_z\bar{G}\bar{G}^\dagger\tilde{D}_z + \bar{\beta}_{2,3}\langle \tilde{z}_{t,3}^\pi, \tilde{z}_t^\pi\rangle \tilde{D}_z\bar{G}\bar{G}^\dagger\tilde{D}_z \\ &\doteq [0, [\tilde{b}_{2,3} - \bar{\beta}_{2,3}]\mathbb{E}\tilde{z}_{t,3}^\Pi(\tilde{z}_{t,3}^\Pi)'\Gamma_{3,2}\Gamma_{3,2}^\dagger] + \bar{\beta}_{2,3}\langle \tilde{z}_{t,3}^\pi, \tilde{z}_t^\pi\rangle \tilde{D}_z\bar{G}\bar{G}^\dagger\tilde{D}_z \\ &= [0, [\tilde{b}_{2,3} - \bar{\beta}_{2,3}]\mathbb{E}\tilde{z}_{t,3}^\Pi(\tilde{z}_{t,3}^\Pi)'\Gamma_{3,2}\Gamma_{3,2}^\dagger] + [0, \bar{\beta}_{2,3}] + \bar{\mathcal{O}}_2\delta G'_{:,2}\left[I - \bar{M}\bar{G}\bar{G}^\dagger\right]\tilde{D}_z \\ &= \tilde{b}_{r,2,:} + [0, [\tilde{b}_{2,3} - \bar{\beta}_{2,3}]\left[\mathbb{E}\tilde{z}_{t,3}^\Pi(\tilde{z}_{t,3}^\Pi)'\Gamma_{3,2}\Gamma_{3,2}^\dagger - I\right]] + \bar{\mathcal{O}}_2\sqrt{T}\delta G'_{:,2}\left[I - \bar{M}\bar{G}\bar{G}^\dagger\right]D_z.\end{aligned}$$

Here the third line follows from the orders of convergence in $\delta G = \bar{G} - \bar{\Gamma}$ established above and $\tilde{b}_{2,3} - \bar{\beta}_{2,3} = o(T^{-\epsilon}), \epsilon > 0$ as follows from standard theory in the stationary case. Then the representation given in (I) is proved by replacing $\bar{\mathcal{O}}_2$ by its limit $\tilde{\mathcal{O}}_2$ which introduces an error of the required form since $\bar{\mathcal{O}}_2 - \tilde{\mathcal{O}}_2 = o(T^{-\epsilon})$ for some $\epsilon > 0$ as follows from $\bar{\beta}_{2,3} - \tilde{b}_{2,3} = o(T^{-\epsilon})$ (see the proof of Lemma A.3).

(II) For (20) note that δG and $\bar{M} - \bar{\Psi}$ both are of order $O((\log T)^3T^{-1/2})$ and the two norm of $\bar{\Psi}$ and $\bar{\Gamma}$ is of order $O(\log T)$. Therefore replacing $\bar{M}\bar{G}\bar{G}^\dagger$ by $\bar{\Psi}\bar{\Gamma}\bar{\Gamma}^\dagger$ introduces an error of order $O((\log T)^4T^{-1/2}) = o(1)$ proving (20) since $\delta G_{3,2} = o(T^{-1/2})$ (see Lemma A.7) and

$$\sqrt{T}\delta G'_{:,2}(I - \bar{\Psi}\bar{\Gamma}\bar{\Gamma}^\dagger) = \sqrt{T}\delta G_{2,2}[-Z_{21}Z_{11}^{-1} \quad I \quad 0] + \sqrt{T}\delta G'_{3,2}[0 \quad 0 \quad I - Z_{33}\bar{\Gamma}_{3,2}\bar{\Gamma}_{3,2}^\dagger].$$

With respect to (19) note that

$$T\delta G'_{:,1}\left[I - \bar{M}\bar{G}\bar{G}^\dagger\right] = T\delta G'_{:,1}\left[I - \bar{\Psi}\bar{\Gamma}\bar{\Gamma}^\dagger\right] + T\delta G'_{:,1}\left[\bar{\Psi}\bar{\Gamma}\bar{\Gamma}^\dagger - \bar{M}\bar{G}\bar{G}^\dagger\right]$$

where

$$T\delta G'_{:,1}\left[I-\bar{\Psi}\bar{\Gamma}\bar{\Gamma}^{\dagger}\right]=\begin{bmatrix}-T\delta G'_{2,1}Z_{21}Z_{11}^{-1} & T\delta G'_{2,1} & T\delta G'_{3,1}\bar{P}\end{bmatrix} \tag{21}$$

is obvious from the form of $I-\bar{\Psi}\bar{\Gamma}\bar{\Gamma}^{\dagger}$ (see the proof of Lemma A.7).

Noting that $\bar{M}\bar{G}\bar{G}^{\dagger}=\bar{M}\bar{G}(\bar{G}'\bar{M}\bar{G})^{-1}\bar{G}'$ it follows that

$$\begin{aligned}\delta G'_{:,1}[\bar{\Psi}\bar{\Gamma}\bar{\Gamma}^{\dagger}-\bar{M}\bar{G}\bar{G}^{\dagger}] &= \delta G'_{:,1}[(I-\bar{\Psi}\bar{\Gamma}\bar{\Gamma}^{\dagger})(\bar{\Psi}-\bar{M})\bar{\Gamma}\bar{\Gamma}^{\dagger}-\bar{\Psi}\bar{\Gamma}(\bar{\Gamma}'\bar{\Psi}\bar{\Gamma})^{-1}\delta G'(I-\bar{\Psi}\bar{\Gamma}\bar{\Gamma}^{\dagger})\\ &\quad -(I-\bar{\Psi}\bar{\Gamma}\bar{\Gamma}^{\dagger})\bar{\Psi}\delta G\bar{\Gamma}^{\dagger}]+o(T^{-1})\\ &= -\delta G'_{3,1}\langle\tilde{z}^{\pi}_{t,3},\tilde{z}^{\pi}_{t,3}\rangle\bar{\Gamma}_{3,2}(\bar{\Gamma}'_{3,2}Z_{3,3}\bar{\Gamma}_{3,2})^{-1}\delta G'_{:,2}(I-\bar{\Psi}\bar{\Gamma}\bar{\Gamma}^{\dagger})+o(T^{-1})\end{aligned}$$

since $\delta G_{1,1}=o(T^{-1}),\delta G_{2,1}=O((\log T)^7/T)$ and (see the proof of Lemma A.7 and (17))

$$\begin{aligned}\sqrt{T}\delta G'_{3,1}(I-Z_{33}\bar{\Gamma}_{3,2}\bar{\Gamma}^{\dagger}_{32}) &= \sqrt{T}[J_{3,1}-\delta^{31}_{z,z}]'\bar{S}(I-Z_{33}\bar{\Gamma}_{3,2}\bar{\Gamma}^{\dagger}_{32})+o(T^{-\epsilon})\\ &= [\langle\tilde{z}^{\pi}_{t,3},\tilde{\varepsilon}_{t,1}\rangle+\langle\tilde{z}^{\pi}_{t,3},\tilde{y}_{t,2}\rangle\Xi']'\bar{S}(I-Z_{33}\bar{\Gamma}_{3,2}\bar{\Gamma}^{\dagger}_{32})+o(T^{-\epsilon})\\ &= o(T^{-\epsilon})\end{aligned}$$

for some $\epsilon>0$ due to $\langle\tilde{z}^{\pi}_{t,3},\tilde{\varepsilon}_{t,1}\rangle=o(T^{-\epsilon})$ for $0<\epsilon<1/2$ and $\langle\tilde{z}^{\pi}_{t,3},\tilde{y}_{t,2}\rangle\to\mathbb{E}\tilde{z}^{\Pi}_{t,3}(\tilde{y}^{\Pi}_{t,2})'=\mathbb{E}\tilde{z}^{\Pi}_{t,3}(\tilde{z}^{\Pi}_{t,3})'\Gamma_{3,2}\tilde{O}'_2$. Now $\langle\tilde{y}^{\pi}_{t,2},\tilde{z}_{t,3}\rangle\bar{S}(I-Z_{33}\bar{\Gamma}_{3,2}\bar{\Gamma}^{\dagger}_{3,2})\to 0$ according to

$$\bar{S}Z_{33}\bar{\Gamma}_{3,2}\to S\mathbb{E}\tilde{z}^{\Pi}_{t,3}(\tilde{z}^{\Pi}_{t,3})'\Gamma_{3,2}. \tag{22}$$

Recall that by definition

$$\mathbb{E}\tilde{z}^{\Pi}_{t,3}(\tilde{z}^{\Pi}_{t,3})'\Gamma_{3,2}\Theta_2^2=\mathbb{E}\tilde{z}^{\Pi}_{t,3}(\tilde{y}^{\Pi}_{t,2})'(\mathbb{E}\tilde{y}^{\Pi}_{t,2}(\tilde{y}^{\Pi}_{t,2})')^{-1}\mathbb{E}\tilde{y}^{\Pi}_{t,2}(\tilde{z}^{\Pi}_{t,3})'\Gamma_{3,2}.$$

Together with the definition of $S=(\mathbb{E}\tilde{z}^{\Pi}_{t,3}(\tilde{z}^{\Pi}_{t,3})'-\mathbb{E}\tilde{z}^{\Pi}_{t,3}(\tilde{y}^{\Pi}_{t,2})'(\mathbb{E}\tilde{y}^{\Pi}_{t,2}(\tilde{y}^{\Pi}_{t,2})')^{-1}\mathbb{E}\tilde{y}^{\Pi}_{t,2}(\tilde{z}^{\Pi}_{t,3})')^{-1}$ this implies

$$S\mathbb{E}\tilde{z}^{\Pi}_{t,3}(\tilde{z}^{\Pi}_{t,3})'\Gamma_{3,2}=\Gamma_{3,2}(I-\Theta_2^2)^{-1}.$$

Finally $\sqrt{T}[\delta^{21}_{yz}-\delta^{21}_{yy}]\langle\tilde{y}^{\pi}_{t,2},\tilde{y}^{\pi}_{t,2}\rangle^{-1}=\Xi+o(T^{-\epsilon})$ is used (as derived above).

Since $\delta G_{3,2}=o(T^{-1/2-\epsilon})$ (see Lemma A.7) it follows from the form of $(I-\bar{\Psi}\bar{\Gamma}\bar{\Gamma}^{\dagger})$ (see the proof of Lemma A.7) that

$$T\delta G'_{:,1}[\bar{\Psi}\bar{\Gamma}\bar{\Gamma}^{\dagger}-\bar{M}\bar{G}\bar{G}^{\dagger}]=-(\sqrt{T}\delta G_{3,1})'\mathbb{E}\tilde{z}^{\Pi}_{t,3}(\tilde{z}^{\Pi}_{t,3})'(\Gamma^{\dagger}_{3,2})'(\sqrt{T}\delta G'_{2,2})[-Z_{21}Z_{11}^{-1},I,0]+o(1). \tag{23}$$

Combining (21) and (23) we obtain

$$T\delta G'_{:,1}\left[I-\bar{M}\bar{G}\bar{G}^{\dagger}\right]=T\left(\delta G'_{2,1}-\delta G'_{3,1}\mathbb{E}\tilde{z}^{\Pi}_{t,3}(\tilde{z}^{\Pi}_{t,3})'(\Gamma^{\dagger}_{3,2})'\delta G'_{2,2}\right)'\begin{bmatrix}-Z_{21}Z_{11}^{-1}, & I, & 0\end{bmatrix}+\begin{bmatrix}0 & 0 & T\delta G'_{3,1}\bar{P}\end{bmatrix}+o(1).$$

This completes the proof by the definition of δH.

A.6 Proof of Lemma A.9

The first claim is standard and follows from Lemma A.1 using $n_t := [\tilde{z}'_{t,1}, (\tilde{z}^u_{t,1})']'$, $v_t := \mathcal{T}_y \Lambda \varepsilon_t$ noting that then v_t is a martingale difference. The second claim is a standard central limit result. Further from Lemma A.7 we have

$$\sqrt{T}\delta G_{2,2} = (T^{-1}\langle \tilde{z}^{\pi}_{t,2.1}, \tilde{z}^{\pi}_{t,2.1}\rangle)^{-1}\sqrt{T}\left[J_{2.1,3}\Gamma_{3,2}\Theta_2^{-2} - \delta_{zz}^{2.1,3}\Gamma_{3,2}\right] + o_P(1).$$

Now from the proof of Lemma A.7

$$\begin{aligned}\sqrt{T}J_{2.1,3}\Gamma_{3,2} &= \langle \tilde{z}^{\pi}_{t,2.1}, \tilde{y}^{\pi}_{t,2}\rangle(\mathbb{E}\tilde{y}^{\Pi}_{t,2}(\tilde{y}^{\Pi}_{t,2})')^{-1}\tilde{O}_2\Gamma'_{3,2}Z_{33}\Gamma_{3,2} + o_P(1)\\ &= \langle \tilde{z}^{\pi}_{t,2.1}, \tilde{\varepsilon}_{t,2}\rangle(\mathbb{E}\tilde{y}^{\Pi}_{t,2}(\tilde{y}^{\Pi}_{t,2})')^{-1}\tilde{O}_2(\tilde{O}_2'(\mathbb{E}\tilde{y}^{\Pi}_{t,2}(\tilde{y}^{\Pi}_{t,2})')^{-1}\tilde{O}_2)^{-1}\Theta_2^2\\ &\quad +\langle \tilde{z}^{\pi}_{t,2.1}, \tilde{z}^{\pi}_{t,3}\rangle\Gamma_{3,2}\Theta_2^2 + o_P(1)\end{aligned}$$

since $\Theta_2^2 = \tilde{O}_2'(\mathbb{E}\tilde{y}^{\Pi}_{t,2}(\tilde{y}^{\Pi}_{t,2})')^{-1}\tilde{O}_2(\Gamma'_{3,2}Z_{33}\Gamma_{3,2})$ as is straightforward to show. This shows the expression for $\sqrt{T}\delta G_{2,2}$ since $\sqrt{T}\delta_{zz}^{2.1,3} = \langle \tilde{z}^{\pi}_{t,2.1}, \tilde{z}^{\pi}_{t,3}\rangle$.
In order to obtain the expression for $T\bar{P}'\delta G_{3,1}$ note that from the definition of $\bar{P}, \bar{P}_{33}$ and the arguments in the proof of Lemma A.7

$$\langle \tilde{z}^{\pi}_{t,3}, \tilde{z}^{\pi}_{t,3}\rangle \bar{P}'\delta G_{3,1} = \bar{P}_{33}\delta G_{3,1} = \bar{P}[J_{3,1} - \delta_{zz}^{31}] + o(T^{-1}).$$

Now $\sqrt{T}[J_{3,1} - \delta_{zz}^{31}] \to \mathbb{E}\tilde{z}^{\Pi}_{t,3}(\tilde{z}^{\Pi}_{t,3})'\Gamma_{3,2}\tilde{O}_2'\Xi'$ according to the proof of Lemma A.7 and $\bar{P} \to I - \mathbb{E}\tilde{z}^{\Pi}_{t,3}(\tilde{z}^{\Pi}_{t,3})'\Gamma_{3,2}\Gamma^{\dagger}_{3,2}$ where the difference is of order $O_P(T^{-1/2})$ since only stationary components are involved. The result then follows from the expression for $J_{3,1}$ according to Lemma A.6 (II) since

$$(I - \mathbb{E}\tilde{z}^{\Pi}_{t,3}(\tilde{z}^{\Pi}_{t,3})'\Gamma_{3,2}\Gamma^{\dagger}_{3,2})\mathbb{E}\tilde{z}^{\Pi}_{t,3}(\tilde{z}^{\Pi}_{t,3})'\Gamma_{3,2}O_2'\Xi' = 0.$$

The evaluations for δH are more involved: Using the expressions given in Lemma A.7 and defining $Z_{22.1} = \langle \tilde{z}_{t,2.1}, \tilde{z}_{t,2.1}\rangle$ we have $\delta H = \delta G_{2,1} - \delta G_{2,2}(\Gamma^{\dagger}_{3,2})'\mathbb{E}\tilde{z}^{\Pi}_{t,3}(\tilde{z}^{\Pi}_{t,3})'\delta G_{3,1} =$

$$\begin{aligned}&= Z_{22.1}^{-1}\left[J_{2.1,1} + \left\{J_{2.1,3} - \delta_{zz}^{2.1,3} - (J_{2.1,3}\Gamma_{3,2}\Theta_2^{-2} - \delta_{zz}^{2.1,3}\Gamma_{3,2})\Gamma_+\right\}\delta G_{3,1}\right] + o(T^{-1})\\ &= Z_{22.1}^{-1}\left[J_{2.1,1} + \left\{J_{2.1,3} - \delta_{zz}^{2.1,3} - \left(J_{2.1,3}\Gamma_{3,2}(\Theta_2^{-2} - I) + (J_{2.1,3} - \delta_{zz}^{2.1,3})\Gamma_{3,2}\right)\Gamma_+\right\}\delta G_{3,1}\right] + o(T^{-1})\\ &= Z_{22.1}^{-1}\left[J_{2.1,1} + (J_{2.1,3} - \delta_{zz}^{2.1,3})\left(I - \Gamma_{3,2}\Gamma_+\right)\delta G_{3,1} - J_{2.1,3}\Gamma_{3,2}(\Theta_2^{-2} - I)(\Gamma^{\dagger}_{3,2})'\mathbb{E}\tilde{z}^{\Pi}_{t,3}(\tilde{z}^{\Pi}_{t,3})'\delta G_{3,1}\right] + o(T^{-1})\\ &= Z_{22.1}^{-1}\left(J_{2.1,1} - J_{2.1,3}/\sqrt{T}\Gamma_{3,2}\Theta_2^{-2}\tilde{O}_2'\Xi'\right) + o_P(T^{-1})\end{aligned}$$

(using $\Gamma_+ := (\Gamma^{\dagger}_{3,2})'\mathbb{E}\tilde{z}^{\Pi}_{t,3}(\tilde{z}^{\Pi}_{t,3})'$) where the last line follows from $T(J_{2.1,3} - \delta_{zz}^{2.1,3})\left(I - \Gamma_{3,2}(\Gamma^{\dagger}_{3,2})'Z_{33}\right)\delta G_{3,1} \to 0$ in probability since $\sqrt{T}[J_{2.1,3} - \delta_{zz}^{2.1,3}]$ converges in distribution, $\sqrt{T}\delta G_{3,1} \to SZ_{33}\Gamma_{3,2}\tilde{O}_2'\Xi'$

and $SZ_{33}\Gamma_{3,2} = \Gamma_{3,2}(I-\Theta_2^2)^{-1}$ (see (22)). The result then follows from some algebraic operations.

The final statement applies Lemma A.3:

$$\bar{\Gamma}_{3,2}' - \Gamma_{3,2}' = \tilde{O}_2^{\dagger}(\langle \tilde{y}_{t,2}^{\pi}, \tilde{z}_{t,3}^{\pi}\rangle(\langle \tilde{z}_{t,3}^{\pi}, \tilde{z}_{t,3}^{\pi}\rangle)^{-1} - \tilde{\beta}_{2,3})(I - S_{p,22}\Gamma_{3,2}') + o(T^{-1/2}). \tag{24}$$

This shows the result since $\bar{\beta}_{2,3} - \tilde{b}_{2,3} = (\bar{\mathcal{O}}_2 - \tilde{O}_2)\bar{\Gamma}_{3,2}' + \tilde{O}_2(\bar{\Gamma}_{3,2}' - \Gamma_{3,2}')$ and $\bar{\Gamma}_{3,2}'(I - Z_{33}\Gamma_{3,2}\Gamma_{3,2}^{\dagger}) \to 0$. The fact that $(I - S_{p,22}\Gamma_{3,2}')(I - Z_{33}\Gamma_{3,2}\Gamma_{3,2}^{\dagger}) = (I - Z_{33}\Gamma_{3,2}\Gamma_{3,2}^{\dagger})$ simplifies the expressions. This concludes the proof.

A.7 Proof of Theorem 3.3

The proof follows closely the proof in the OLS case. The changes in comparison to the OLS case are that $\langle y_t, z_t\rangle$ is replaced with

$$\hat{\Sigma}_{y,z} := \langle y_t, z_t^{\pi}\rangle - \hat{\Delta}_{\hat{u},\Delta z^{\pi}} - \hat{\Omega}_{\hat{u},\Delta z}\hat{\Omega}_{\Delta z,\Delta z}^{-1}(\langle \Delta z_t, z_t^{\pi}\rangle - \hat{\Delta}_{\Delta z,\Delta z^{\pi}}) = \langle y_t, z_t^{\pi}\rangle + \hat{B}^{+}$$

where hence the additional term is called $\hat{B}^{+}$ and in the SVD the weighting is not based on $\langle y_t^{\pi}, y_t^{\pi}\rangle$ but on $W_{+} = \left(\langle y_t^{\pi}, y_t^{\pi}\rangle + \hat{C}^{+}\right)^{-1}$ where

$$\hat{C}^{+} := -\hat{\Omega}_{\hat{u},\Delta z}\hat{\Omega}_{\Delta z,\Delta z}^{-1}(\langle \Delta z_t, y_t^{\pi}\rangle - \hat{\Delta}_{\Delta z,\Delta y^{\pi}}) - (\langle \Delta z_t, y_t^{\pi}\rangle - \hat{\Delta}_{\Delta z,\Delta y^{\pi}})'(\hat{\Omega}_{\hat{u},\Delta z}\hat{\Omega}_{\Delta z,\Delta z}^{-1})'.$$

The asymptotics for the additional terms are detailed in the lemma below:

Lemma A.10 *Under the assumptions of Theorem 3.3 the following holds:*

$$\begin{aligned}
\hat{\Delta}_{\hat{u},\Delta \tilde{z}^{\pi}} + \hat{\Omega}_{\hat{u},\Delta z}\hat{\Omega}_{\Delta z,\Delta z}^{-1}(\langle \Delta z_t, \tilde{z}_t^{\pi}\rangle - \hat{\Delta}_{\Delta z,\Delta \tilde{z}^{\pi}}) &= \left[\ \Omega_{\hat{u}\Delta z}^{:,n}(\Omega_{\Delta z\Delta z}^{n,n})^{-1}\int dB_n B_n' + o_P(1) \quad o_P(T^{-1/2})\ \right], \\
\hat{\Omega}_{\hat{u},\Delta z}\hat{\Omega}_{\Delta z,\Delta z}^{-1}(\langle \Delta z_t, \tilde{y}_t^{\pi}\rangle - \hat{\Delta}_{\Delta z,\Delta \tilde{y}^{\pi}}) &= \left[\ \Omega_{\hat{u}\Delta z}^{:,n}(\Omega_{\Delta z\Delta z}^{n,n})^{-1}\int dB_n B_n' \begin{pmatrix} I \\ 0 \end{pmatrix} + o_P(1) \quad o_P(T^{-1/2})\ \right]
\end{aligned}$$

Proof: The proof of the first statement uses the fact that according to Lemma A.2 $\hat{\Delta}_{\hat{u},\Delta z} = [o_P(1), o_P(T^{-1/2})]$. Here the restrictions on the increase of K as a function of T is used such that we obtain $\sqrt{K/T} = o_P(1), 1/\sqrt{KT} = o_P(T^{-1/2}), K^{-2} = o_P(T^{-1/2})$. Hence it is of lower order compared to the leading terms. Further $\hat{\Omega}_{\hat{u},\Delta z}\hat{\Omega}_{\Delta z,\Delta z}^{-1}(\langle \Delta z_t, \tilde{z}_t\rangle - \hat{\Delta}_{\Delta z,\Delta \tilde{z}}) = [O_P(1), o_P(T^{-1/2})]$ and hence these terms are of the same order as the leading terms. The proof of the second statement is an easy consequence of the results listed in Lemma A.2 using $\tilde{y}_t^{\pi} = \tilde{b}_r\tilde{z}_t^{\pi} + \tilde{\varepsilon}_t^{\pi}$. Note that

$$\langle \Delta z_t, \tilde{y}_t^{\pi}\rangle - \hat{\Delta}_{\Delta z,\Delta \tilde{y}^{\pi}} = (\langle \Delta z_t, \tilde{z}_t^{\pi}\rangle - \hat{\Delta}_{\Delta z,\Delta \tilde{z}^{\pi}})\tilde{b}_r' + (\langle \Delta z_t, \tilde{u}_t\rangle - \hat{\Delta}_{\Delta z,\Delta \tilde{u}^{\pi}})$$

Convergence for the first summand is contained as the first statement in the lemma, while again following Lemma A.2 we obtain convergence for the second term. □

For both $\hat{\Sigma}_{y,z}$ and W_+ after transformation using the matrices $\mathcal{T}_y, \mathcal{T}_z$ the additional terms in the diagonal blocks are of lower order than the original terms. For the off-diagonal blocks the additional terms are of the same order in probability. This follows from the results in Lemma A.2. However, for the off-diagonal terms in the consistency proof only the order of convergence is used. Consequently the consistency result and the order of convergence (in probability) also hold in the FM case.

In the following we will use the following definitions using the same notation as in the OLS case in order to avoid the introduction of new symbols.

$$
\begin{aligned}
\bar{Q} &:= \hat{\Sigma}'_{y,z} W_+^{-1} \hat{\Sigma}_{y,z}, \\
\bar{M} &:= \langle \tilde{D}_z \tilde{z}_t^\pi, \tilde{D}_z \tilde{z}_t^\pi \rangle, \\
\bar{\Phi} &:= \begin{bmatrix} T^{-1}\langle \tilde{z}_{t,1}^\pi, \tilde{z}_{t,1}^\pi \rangle & T^{-1}\langle \tilde{z}_{t,1}^\pi, \tilde{z}_{t,2}^\pi \rangle & 0 \\ T^{-1}\langle \tilde{z}_{t,2}^\pi, \tilde{z}_{t,1}^\pi \rangle & T^{-1}\langle \tilde{z}_{t,2}^\pi, \tilde{z}_{t,1}^\pi \rangle \langle \tilde{z}_{t,1}^\pi, \tilde{z}_{t,1}^\pi \rangle^{-1} \langle \tilde{z}_{t,1}^\pi, \tilde{z}_{t,2}^\pi \rangle & 0 \\ 0 & 0 & \langle \tilde{z}_{t,3}^\pi, \tilde{y}_{t,2}^\pi \rangle \hat{\Sigma}_{y2,y2}^{-1} \langle \tilde{y}_{t,2}^\pi, \tilde{z}_{t,3}^\pi \rangle \end{bmatrix}, \\
\bar{\Psi} &:= \begin{bmatrix} T^{-1}\langle \tilde{z}_{t,1}^\pi, \tilde{z}_{t,1}^\pi \rangle & T^{-1}\langle \tilde{z}_{t,1}^\pi, \tilde{z}_{t,2}^\pi \rangle & 0 \\ T^{-1}\langle \tilde{z}_{t,2}^\pi, \tilde{z}_{t,1}^\pi \rangle & T^{-1}\langle \tilde{z}_{t,2}^\pi, \tilde{z}_{t,2}^\pi \rangle & 0 \\ 0 & 0 & \langle \tilde{z}_{t,3}^\pi, \tilde{z}_{t,3}^\pi \rangle \end{bmatrix}.
\end{aligned}
$$

Here

$$
\hat{\Sigma}_{y2,y2} := \langle \tilde{y}_{t,2}^\pi, \tilde{y}_{t,2}^\pi \rangle - [0, I]' \left[\hat{\Omega}_{\hat{u},\Delta z} \hat{\Omega}_{\Delta z,\Delta z}^{-1} (\langle \Delta z_t, \tilde{y}_t^\pi \rangle - \hat{\Delta}_{\Delta z, \Delta \tilde{y}}) - (\langle \Delta z_t, \tilde{y}_t^\pi \rangle - \hat{\Delta}_{\Delta z, \Delta \tilde{y}})' (\hat{\Omega}_{\hat{u},\Delta z} \hat{\Omega}_{\Delta z,\Delta z}^{-1})' \right] [0, I]'
$$

such that $\hat{\Sigma}_{y2,y2} \to \Sigma_{y2,y2} := \mathbb{E} \tilde{y}_{t,2}^\Pi (\tilde{y}_{t,2}^\Pi)'$.

Hence the definition of $\bar{Q}$ is adapted to the SVD occurring in the FM estimation. Also the $(3,3)$ entry of $\bar{\Phi}$ is changed slightly. The reason for this is visible in the next Lemma A.11 which is the analogon to Lemma A.6 for the OLS case.

Lemma A.11 *Let the assumptions of Theorem 3.3 hold.*

(I) Partition the matrices $\bar{Q}, \bar{M}, \bar{\Phi}, \bar{\Psi}$ according to the partitioning of $\tilde{z}_t$ denoting the various blocks using subscripts. Then:

$$
\begin{aligned}
\delta_{zz} &:= \bar{M} - \bar{\Psi} = \begin{bmatrix} 0 & 0 & O_P(T^{-1/2}) \\ 0 & 0 & O_P(T^{-1/2}) \\ O_P(T^{-1/2}) & O_P(T^{-1/2}) & 0 \end{bmatrix}, \\
\delta_{yz} &:= \begin{bmatrix} T^{-1}(\langle \tilde{y}_{t,1}^\pi, \tilde{z}_{t,1}^\pi \rangle - \langle \tilde{z}_{t,1}^\pi, \tilde{z}_{t,1}^\pi \rangle) & T^{-1}(\langle \tilde{y}_{t,1}^\pi, \tilde{z}_{t,2}^\pi \rangle - \langle \tilde{z}_{t,1}^\pi, \tilde{z}_{t,2}^\pi \rangle) & T^{-1/2}\langle \tilde{y}_{t,1}^\pi, \tilde{z}_{t,3}^\pi \rangle \\ T^{-1/2}\langle \tilde{y}_{t,2}^\pi, \tilde{z}_{t,1}^\pi \rangle & T^{-1/2}\langle \tilde{y}_{t,2}^\pi, \tilde{z}_{t,2}^\pi \rangle & 0 \end{bmatrix} + \tilde{D}_y \tilde{B}^+ \tilde{D}_z
\end{aligned}
$$

$$
\begin{aligned}
&= \begin{bmatrix} O_P(T^{-1}) & O_P(T^{-1}) & O_P(T^{-1/2}) \\ O_P(T^{-1/2}) & O_P(T^{-1/2}) & o_P(1) \end{bmatrix}, \\
\delta_{yy} &:= \tilde{D}_y \left(\langle \tilde{y}_t^\pi, \tilde{y}_t^\pi \rangle + \tilde{C}^+ - \begin{bmatrix} \langle z_{t,1}^\pi, z_{t,1}^\pi \rangle & 0 \\ 0 & \hat{\Sigma}_{y2,y2} \end{bmatrix} \right) \tilde{D}_y \\
&= \begin{bmatrix} O_P(T^{-1}) & O_P(T^{-1/2}) \\ O_P(T^{-1/2}) & 0 \end{bmatrix}.
\end{aligned}
$$

(II) Let $J := \bar{Q} - \bar{\Phi}$. To simplify notation define $Z_{ij} := T^{-1}\langle \tilde{z}_{t,i}^\pi, \tilde{z}_{t,j}^\pi \rangle, i,j = 1,2$. Then

$$
\begin{aligned}
J_{i,j} &= [\delta_{zy}^{i1} - Z_{i1} Z_{11}^{-1} \delta_{yy}^{11}] Z_{11}^{-1} Z_{1j} + Z_{i1} Z_{11}^{-1} \delta_{yz}^{1j} \\
&\quad + [\delta_{zy}^{i2} - Z_{i1} Z_{11}^{-1} \delta_{yy}^{12}] \Sigma_{y2,y2}^{-1} [\delta_{yz}^{2j} - \delta_{yy}^{21} Z_{11}^{-1} Z_{1j}] + o_P(T^{-1}), \\
J_{3,i} &= \delta_{zy}^{31} Z_{11}^{-1} Z_{1i} + \langle \tilde{z}_{t,3}^\pi, \tilde{y}_{t,2}^\pi \rangle \hat{\Sigma}_{y2,y2}^{-1} [\delta_{yz}^{2i} - \delta_{yy}^{21} Z_{11}^{-1} Z_{1i}] + o_P(T^{-1}), \\
J_{3,3} &= \left[\langle \tilde{z}_{t,3}^\pi, \tilde{y}_{t,2}^\pi \rangle \hat{\Sigma}_{y2,y2}^{-1} \delta_{yy}^{21} - \delta_{zy}^{31} \right] Z_{11}^{-1} \left[\delta_{yy}^{12} \hat{\Sigma}_{y2,y2}^{-1} \langle \tilde{y}_{t,2}^\pi, \tilde{z}_{t,3}^\pi \rangle - \delta_{yz}^{13} \right] + o_P(T^{-1})
\end{aligned} \tag{25}
$$

for $i = 1,2, j = 1,2$ where expressions for the remaining blocks of J follow from symmetry. Hence $J_{i,j} = O_P(T^{-1})$ for $i,j = 1,2$. Further $J_{3,i} = O_P(T^{-1/2})$ for $i = 1,2,3$. $J_{3,3} = O_P(T^{-1})$ and $J_{3,3} = O((\log T)^3/T)$ respectively.

(III) $\delta G = O_P(T^{-1/2})$.

PROOF: (I) follows from Lemma A.10. Note that compared to the OLS case in the $(3,3)$ entry of $\bar{\Phi}$ the matrix $\langle \tilde{y}_{t,2}^\pi, \tilde{y}_{t,2}^\pi \rangle$ is replaced with $\hat{\Sigma}_{y2,y2} = \langle \tilde{y}_{t,2}^\pi, \tilde{y}_{t,2}^\pi \rangle + o_P(T^{-1/2})$ in order to obtain $\delta_{yy}^{22} = 0$ rather than $o_P(T^{-1/2})$. Only the in probability statements are used.

The proof of (II) then is unchanged except that $\delta_{yz}^{2,3} = o_P(T^{-1/2})$ needs to be taken into account. (III) is then immediate. □

Next the proof of Lemma A.7 uses only the results of Lemma A.6 and equation (12) in combination with the following limit results:

$$
\begin{aligned}
\sqrt{T}[\delta_{yz}^{21} - \delta_{yy}^{21}]' \langle \tilde{y}_{t,2}^\pi, \tilde{y}_{t,2}^\pi \rangle^{-1} &= \Xi + o(1), \\
\sqrt{T}[J_{3,1} - \delta_{zz}^{3,1}] &= \mathbb{E} \tilde{z}_{t,3}^\Pi (\tilde{y}_{t,2}^\Pi)' \Xi' + o_P(1), \\
T \langle \tilde{z}_{t,2.1}^\pi, \tilde{z}_{t,2.1}^\pi \rangle^{-1} J_{2.1,1} &= (T^{-1} \langle \tilde{z}_{t,2.1}^\pi, \tilde{z}_{t,2.1}^\pi \rangle)^{-1} \left(\langle \tilde{z}_{t,2.1}^\pi, \tilde{\varepsilon}_{t,1.2} \rangle) + \tilde{B}_{1,2.1}' + \tilde{B}_{2,2.1}' \Xi' \right)' + o_P(1), \\
\sqrt{T} \langle \tilde{z}_{t,2.1}^\pi, \tilde{z}_{t,2.1}^\pi \rangle^{-1} J_{2.1,3} &= \langle \tilde{z}_{t,2.1}^\pi, \tilde{z}_{t,2.1}^\pi \rangle^{-1} \left(\langle \tilde{z}_{t,2.1}^\pi, \tilde{y}_{t,2}^\pi + \rangle + \tilde{B}_{2.1,2} \right)' \Sigma_{y2,y2}^{-1} \tilde{O}_2 \Gamma_{3,2}' Z_{33} + o_P(1).
\end{aligned}
$$

where again $\Xi := -\mathbb{E} \tilde{\varepsilon}_{t,1} \tilde{y}_{t,2}' \Sigma_{y2,y2}^{-1}$ and $\tilde{\varepsilon}_{t,1.2} = \tilde{\varepsilon}_{t,1} + \Xi \tilde{y}_{t,2}^\Pi$. Further ($\tilde{B}^+$ denoting the transformed quantity $\hat{B}^+$)

$$
\tilde{B}_{1,2.1} = [I, 0] \tilde{B}^+ [-Z_{21} Z_{11}^{-1}, I, 0]', \quad \tilde{B}_{2,2.1} = [0, I] \tilde{B}^+ [-Z_{21} Z_{11}^{-1}, I, 0]'.
$$

Here the first statement follows from Lemma A.10. The second from the fact that the (1,3) block of $\tilde{\Sigma}_{y,z}$ is of order $o_P(T^{-1/2})$ relating to a stationary component of the regressors. The remaining statements follow straightforwardly from the definition of J.

Lemma A.8 needs to be changed slightly by replacing a.s. statements by the corresponding in probability version.

Lemma A.12 *Let the assumptions of Theorem 3.3 hold.*

(I) Then

$$\mathcal{T}_y(\hat{\beta}^+_{RRR,r} - b_r)\mathcal{T}_z^{-1}D_z^{-1} = \sqrt{T}\left[\langle\tilde{\varepsilon}_t, \tilde{z}_t^\pi\rangle + \tilde{B}^+\right]\tilde{D}_z\bar{\Gamma}\bar{\Gamma}^\dagger + \begin{bmatrix} TI & 0 \\ 0 & \sqrt{T}O \end{bmatrix}\delta G'(I - \bar{M}\bar{G}\bar{G}^\dagger)$$
$$+ \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \sqrt{T}[\tilde{\beta}_{2,3} - \tilde{b}_{2,3}][I - \mathbb{E}\tilde{z}^\Pi_{t,3}(\tilde{z}^\Pi_{t,3})'\Gamma_{3,2}\Gamma^\dagger_{3,2}] \end{bmatrix} + o_P(1)$$

where

$$\bar{\beta}_{2,3} = \langle\tilde{y}^\pi_{t,2}, \tilde{z}^\pi_{t,3}\rangle\bar{\Gamma}_{3,2}\bar{\Gamma}^\dagger_{3,2} = \bar{O}_2\bar{\Gamma}'_{3,2} \to \tilde{O}_2\Gamma'_{3,2} = \tilde{b}_{2,3}$$

denotes the solution to the subproblem of the problem (b) corresponding to the stationary components. $\bar{\Gamma}^\dagger_{3,2} := (\bar{\Gamma}'_{3,2}\langle\tilde{z}^\pi_{t,3}, \tilde{z}^\pi_{t,3}\rangle\bar{\Gamma}_{3,2})^{-1}\bar{\Gamma}'_{3,2}$.

(II) Letting $\delta G_{:,1}$ and $\delta G_{:,2}$ denote the first and second block column of δG it holds that

$$T\delta G'_{:,1}(I - \bar{M}\bar{G}\bar{G}^\dagger) = \left[-T\delta H' Z_{21}Z_{11}^{-1} \quad T\delta H' \quad T\delta G'_{3,1}\bar{P}\right] + o_P(1), \tag{26}$$
$$\sqrt{T}\delta G'_{:,2}(I - \bar{M}\bar{G}\bar{G}^\dagger) = \sqrt{T}\delta G'_{2,2}\left[-Z_{21}Z_{11}^{-1} \quad I \quad 0\right] + o_P(1) \tag{27}$$

where $\delta H = \delta G_{2,1} - \delta G_{2,2}(\Gamma^\dagger_{3,2})'\mathbb{E}\tilde{z}^\Pi_{t,3}(\tilde{z}^\Pi_{t,3})'\delta G_{3,1}$ *and* $\bar{P} = I - Z_{33}\bar{\Gamma}_{3,2}\bar{\Gamma}^\dagger_{3,2}$.

The only change in the proof consists in exchanging the estimation error for the OLS estimator by the estimation error for the FM estimator in (I). The rest of the proof is analogously to the OLS case and hence omitted. Primarily the orders of convergence derived above are used.

It remains to analyze the asymptotic distribution of the various terms. This is done in the analogon to Lemma A.9:

Lemma A.13 *With* $\tilde{O}_2^\dagger = (\tilde{O}_2'(\mathbb{E}\tilde{y}^\Pi_{t,2}(\tilde{y}^\Pi_{t,2})')^{-1}\tilde{O}_2)^{-1}\tilde{O}_2'(\mathbb{E}\tilde{y}^\Pi_{t,2}(\tilde{y}^\Pi_{t,2})')^{-1}$ *we have*

$$(\langle\tilde{\varepsilon}_t, \tilde{z}^\pi_{t,1}\rangle + \tilde{B}^+_{:,1})Z_{11}^{-1} \xrightarrow{d} f(\mathcal{T}_y\Lambda B, W^\Pi_{z,1}, 0),$$
$$\sqrt{T}\langle\tilde{\varepsilon}_t, \tilde{z}^\pi_{t,3}\rangle \xrightarrow{d} \mathcal{N}(0, V),$$
$$\sqrt{T}\delta G_{2,2} = Z_{22.1}^{-1}(\langle\tilde{\varepsilon}^\pi_{t,2}, \tilde{z}^\pi_{t,2.1}\rangle + \tilde{B}_{2,2.1})'(\tilde{O}_2^\dagger)' + o_P(1) \xrightarrow{d} M'_{2,+}(\tilde{O}_2^\dagger)',$$

$$
\begin{aligned}
T\bar{P}'\delta G_{3,1} &= Z_{33}^{-1}P'\sqrt{T}\langle \tilde{z}_{t,3}^{\pi}, \tilde{\varepsilon}_{t,1.2}\rangle + Z_{33}^{-1}\sqrt{T}(\bar{P}-P)'\mathbb{E}\tilde{z}_{t,3}^{\Pi}(\tilde{y}_{t,2}^{\Pi})'\Xi' + o_P(1) \to \tilde{R}', \\
T\delta H &= Z_{22.1}^{-1}\left[\langle \tilde{z}_{t,2.1}^{\pi}, \tilde{\varepsilon}_{t,1}\rangle + \tilde{B}_{1,2.1}' + \left(\langle \tilde{z}_{t,2.1}^{\pi}, \tilde{\varepsilon}_{t,2}^{\pi}\rangle + \tilde{B}_{2,2.1}'\right)(I - \tilde{O}_2\tilde{O}_2^{\dagger})'\Xi'\right] + o_P(1) \\
&\xrightarrow{d} N_{+}', \\
\sqrt{T}[\bar{\beta}_{2,3} - \tilde{\beta}_{2,3}]P &= \tilde{O}_2\tilde{O}_2^{\dagger}\sqrt{T}\langle \tilde{\varepsilon}_{t,2}, \tilde{z}_{t,3}^{\Pi}\rangle(\mathbb{E}\tilde{z}_{t,3}^{\Pi}(\tilde{z}_{t,3}^{\Pi})')^{-1}P + o_P(1),
\end{aligned}
$$

where $P = (I - \mathbb{E}\tilde{z}_{t,3}^{\Pi}(\tilde{z}_{t,3}^{\Pi})'\Gamma_{3,2}\Gamma_{3,2}^{\dagger})$. *Here* $M_{2,+} = f([0,I]\mathcal{T}_y B, W_{z,2}^{\Pi} - Y_{21}Y_{11}^{-1}W_{z,1}^{\Pi}, 0)$ *and* $N_{+} = f([[I,0] + \Xi(I - O_2O_2^{\dagger})[0,I]]\mathcal{T}_y B, W_{z,2}^{\Pi} - Y_{21}Y_{11}^{-1}W_{z,1}^{\Pi}, 0)$. $T\bar{P}\delta G_{3,1}$ *and* $\sqrt{T}[\tilde{\beta}_{2,3} - \beta_2]P$ *converge in distribution to Gaussian random variables with mean zero.*

The proof follows analogously to the OLS case. The remaining steps of the proof are analogous to the proof for Theorem 3.2 and hence omitted.

B Collection of notation

In this section the notation is presented in order to make reference easier. The general concept is to use lower case letters for processes (where y is reserved for the dependent variable, z denotes regressors and v, w, u is reserved for stationary processes, ε, η denote white noise). Processes built using a number of coordinates of other processes are indicated using sub- or superscripts. Regression residuals are indicated using a superscript π (where the regressors are clear from the context) and their corresponding limits with a superscript Π (this notation is only used, if limits exist).

Upper case letters are used for matrices. Matrices that transform the basis of processes are indicated using $\mathcal{T}$ where the transformed process is indicated as a subscript. Scaling matrices that are introduced in order to ensure the convergence of matrices are denoted using D with subscripts denoting the processes to which they are applied.

Estimates in the original basis are denoted using a $\hat{\bullet}$, in the transformed basis (see Theorem 3.1) with a $\tilde{\bullet}$ and in the transformed basis with appropriate scaling ensuring convergence with a $\bar{\bullet}$ or $\check{\bullet}$ respectively.

B.1 Processes

Below a_t, b_t are used to denote arbitrary processes, where the notation applies to a number of different processes.

$$
\begin{aligned}
y_t &= b_r z_t^r + b_u z_t^u + L\varepsilon_t, \\
z_t &= \begin{bmatrix} z_t^r \\ z_t^u \end{bmatrix}, \mathrm{diag}(\Delta, I) H_r' z_t^r = v_t, \mathrm{diag}(\Delta, I) H_u' z_t^u = w_t, \\
\nu_t &= \begin{bmatrix} v_t \\ w_t \end{bmatrix} = c(z)\varepsilon_t, c(0) = 0, \det c(1) \neq 0, \\
\langle a_t, b_t \rangle &= T^{-1} \sum_{t=1}^{T} a_t b_t', \\
a_t^\pi &= a_t - \langle a_t, z_t^u \rangle \langle z_t^u, z_t^u \rangle^{-1} z_t^u \left(\to a_t^\Pi (\text{if convergent}) \right), \\
\tilde{y}_t &= \mathcal{T}_y (y_t - b_u z_t^u) = \tilde{b}_r \tilde{z}_t + \tilde{\varepsilon}_t = \begin{bmatrix} I & 0 & 0 \\ 0 & 0 & \tilde{b}_{2,3} \end{bmatrix} \begin{bmatrix} \tilde{z}_{t,1} \\ \tilde{z}_{t,2} \\ \tilde{z}_{t,3} \end{bmatrix} + \begin{bmatrix} \tilde{\varepsilon}_{t,1} \\ \tilde{\varepsilon}_{t,2} \end{bmatrix}, \\
\tilde{z}_t &= \mathcal{T}_{z,r} z_t^r = \begin{bmatrix} \tilde{z}_{t,1} \\ \tilde{z}_{t,2} \\ \tilde{z}_{t,3} \end{bmatrix}, \\
\Delta \tilde{z}_{t,1} &= \tilde{c}_{z,1}(z)\varepsilon_t, \Delta \tilde{z}_{t,2} = \tilde{c}_{z,2}(z)\varepsilon_t, \tilde{z}_{t,3} = \tilde{c}_{z,3}(z)\varepsilon_t \quad \text{stationary}
\end{aligned}
$$

$$
\begin{aligned}
\tilde{z}_t^u &= \mathcal{T}_{z,u} z_t^u = \begin{bmatrix} \tilde{z}_{t,1}^u \\ \tilde{z}_{t,2}^u \end{bmatrix}, \\
\Delta \tilde{z}_{t,1}^u &= \tilde{c}_{u,1}(z)\varepsilon_t, \tilde{z}_{t,2}^u = \tilde{c}_{u,2}(z)\varepsilon_t \quad \text{stationary}, \\
\begin{bmatrix} \tilde{c}_{z,1}(1) \\ \tilde{c}_{z,2}(1) \\ \tilde{c}_{u,1}(1) \end{bmatrix} & \quad \text{is of full row rank.}
\end{aligned}
$$

B.2 Matrices

b_r $\in \mathbb{R}^{s \times m_r}$ coefficient matrix corresponding to z_t^r

b_u $\in \mathbb{R}^{s \times m_u}$ coefficient matrix corresponding to z_t^u

b $= [b_r, b_u]$ coefficient matrix corresponding to z_t.

$\mathcal{T}_y$ $\in \mathbb{R}^{s \times s}$ used to transform y_t into $\tilde{y}_t$ separating stationary from nonstationary terms.

$\mathcal{T}_{z,r}$ $\in \mathbb{R}^{m_r \times m_r}$ used to transform z_t^r into $\tilde{z}_t$ separating stationary from nonstationary terms.

$\mathcal{T}_{z,u}$ $\in \mathbb{R}^{m_u \times m_u}$ used to transform z_t^u into $\tilde{z}_t^u$ separating stationary from nonstationary terms.

$\mathcal{T}_z$ $= \text{diag}(\mathcal{T}_{z,r}, \mathcal{T}_{z,u})$.

$\tilde{b}_r$ $= \mathcal{T}_y b_r \mathcal{T}_{z,r}^{-1}$.

$\tilde{b}_u$ $= \mathcal{T}_y b_u \mathcal{T}_{z,u}^{-1}$.

$\hat{\beta}_{OLS}$ OLS estimator of β.

$\hat{\beta}_{OLS,r}$ OLS estimator of b_r

$\hat{\beta}_{OLS,u}$ OLS estimator of b_u

$\tilde{\beta}_{OLS}$ OLS estimator of $\tilde{b}$.

$\tilde{\beta}_{OLS,r}$ OLS estimator of $\tilde{b}_r$

$\tilde{\beta}_{OLS,u}$ OLS estimator of $\tilde{b}_u$

$\hat{\beta}_{RRR}$ RRR estimator of b.

$\hat{\beta}_{RRR,r}$ RRR estimator of b_r

$\hat{\beta}_{RRR,u}$ RRR estimator of b_u

$\tilde{\beta}_{RRR}$ RRR estimator of $\tilde{b}$.

$\tilde{\beta}_{RRR,r}$ RRR estimator of $\tilde{b}_r$

$\tilde{\beta}_{RRR,u}$ RRR estimator of $\tilde{b}_u$

$\hat{\Xi}_+$ $\in \mathbb{R}^{s\times s}$ weighting matrix.

Ξ $= -\mathbb{E}\tilde{\varepsilon}_{t,1}\tilde{y}'_{t,2}(\mathbb{E}\tilde{y}^{\Pi}_{t,2}\tilde{y}'_{t,2})^{-1}$

M_r $= f(W, W_z^{\Pi})$ where W denotes the Brownian motion corresponding to $(\varepsilon_t)_{t\in\mathbb{N}}$, $W_z = \tilde{c}_{1:2}(1)W, W_u = \tilde{c}_{u,1}(1)W, W_z^{\Pi} = W_z - \int W_z W_u'(\int W_u W_u')^{-1}W_u$. Further $f(W_1, W_2) = \int dW_1 W_2(\int W_2 W_2')^{-1}$.

M_u $= f(W, W_u)$.

N_r $= \int W_z W_u'(\int W_u W_u')^{-1}$.

$M_{r,2}$ $= f([0, I]\mathcal{T}_y W, W^{\Pi}_{z,2} - Y_{21}Y_{11}^{-1}W^{\Pi}_{z,1})$.

Y_{i1} $= \int W^{\Pi}_{z,i}(W^{\Pi}_{z,1})'$.

P $= I - \mathbb{E}\tilde{z}^{\Pi}_{t,3}(\tilde{z}^{\Pi}_{t,3})'\Gamma_{3,2}\Gamma^{\dagger}_{3,2}, \Gamma^{\dagger}_{3,2} = (\Gamma'_{3,2}\mathbb{E}\tilde{z}^{\Pi}_{t,3}(\tilde{z}^{\Pi}_{t,3})'\Gamma_{3,2})^{-1}\Gamma_{3,2}$.

$\bar{P}_{3,3}$ $= \langle \tilde{z}^{\pi}_{t,3}, \tilde{z}^{\pi}_{t,3}\rangle - \langle \tilde{z}^{\pi}_{t,3}, \tilde{z}^{\pi}_{t,3}\rangle \bar{\Gamma}_{3,2}\bar{\Gamma}^{\dagger}_{3,2}\langle \tilde{z}^{\pi}_{t,3}, \tilde{z}^{\pi}_{t,3}\rangle$.

D_y $= \mathrm{diag}(T^{-1}I_{c_y}, T^{-1/2})$ proper scaling for $\tilde{y}^{\pi}_t$.

$D_{z,r}$ $= \mathrm{diag}(T^{-1}I_{c_r}, T^{-1/2})$ proper scaling for $\tilde{z}^{\pi}_t$.

$D_{z,u}$ $= \mathrm{diag}(T^{-1}I_{c_u}, T^{-1/2})$ proper scaling for $\tilde{z}^{u}_t$.

D_z $= \mathrm{diag}(D_{z,r}, D_{z,u})$

$\tilde{D}_z$ $= D_z T^{1/2} = \mathrm{diag}(T^{-1/2}I, I), \quad \tilde{D}_y = D_y T^{1/2} = \mathrm{diag}(T^{-1/2}I, I)$

$\bar{G}$ $= \tilde{D}_z^{-1}\hat{G}$

$\bar{Q}$ $= \langle \tilde{D}_z\tilde{z}^{\pi}_t, \tilde{D}_y\tilde{y}^{\pi}_t\rangle\langle \tilde{D}_y\tilde{y}^{\pi}_t, \tilde{D}_y\tilde{y}^{\pi}_t\rangle^{-1}\langle \tilde{D}_y\tilde{y}^{\pi}_t, \tilde{D}_z\tilde{z}^{\pi}_t\rangle$

$\bar{M}$ $= \langle \tilde{D}_z\tilde{z}^{\pi}_t, \tilde{D}_z\tilde{z}^{\pi}_t\rangle$

$$\bar{\Phi} = \begin{bmatrix} T^{-1}\langle \tilde{z}^{\pi}_{t,1}, \tilde{z}^{\pi}_{t,1}\rangle & T^{-1}\langle \tilde{z}^{\pi}_{t,1}, \tilde{z}^{\pi}_{t,2}\rangle & 0 \\ T^{-1}\langle \tilde{z}^{\pi}_{t,2}, \tilde{z}^{\pi}_{t,1}\rangle & T^{-1}\langle \tilde{z}^{\pi}_{t,2}, \tilde{z}^{\pi}_{t,1}\rangle\langle \tilde{z}^{\pi}_{t,1}, \tilde{z}^{\pi}_{t,1}\rangle^{-1}\langle \tilde{z}^{\pi}_{t,1}, \tilde{z}^{\pi}_{t,2}\rangle & 0 \\ 0 & 0 & \langle \tilde{z}^{\pi}_{t,3}, \tilde{y}^{\pi}_{t,2}\rangle\langle \tilde{y}^{\pi}_{t,2}, \tilde{y}^{\pi}_{t,2}\rangle^{-1}\langle \tilde{y}^{\pi}_{t,2}, \tilde{z}^{\pi}_{t,3}\rangle \end{bmatrix}$$

Original	Formula	$\langle z_t^{r,\pi}, y_t^\pi\rangle\langle y_t^\pi, y_t^\pi\rangle^{-1}\langle y_t^\pi, z_t^{r,\pi}\rangle\hat{G} = \langle z_t^{r,\pi}, z_t^{r,\pi}\rangle\hat{G}\hat{R}^2$
Transformed	Formula	$\langle \tilde{z}_t^{\pi}, \tilde{y}_t^\pi\rangle\langle \tilde{y}_t^\pi, \tilde{y}_t^\pi\rangle^{-1}\langle \tilde{y}_t^\pi, \tilde{z}_t^{\pi}\rangle\tilde{G} = \langle \tilde{z}_t^{\pi}, \tilde{z}_t^{\pi}\rangle\tilde{G}\tilde{R}^2$
	Relations	$\tilde{G} = \mathcal{T}_{z,r}^{-1}\hat{G}$
Scaled	Formula	$\bar{Q}\bar{G} = \bar{M}\bar{G}\bar{R}^2$
	Restrictions	$[I, 0]\bar{G}_{:,1} = I, S'_{p,2}\bar{G}_{:,2} = I, \bar{R}^2 = \mathrm{diag}(\bar{R}_1^2, \bar{R}_2^2)$
	Relations	$\bar{G} = \tilde{D}_z^{-1}\tilde{G}$
Decoupled	Formula	$\bar{\Phi}\bar{\Gamma} = \bar{\Psi}\bar{\Gamma}\bar{\Theta}^2$
	Restrictions	$\bar{\Gamma}' S_p = I, \bar{\Theta}^2 = \mathrm{diag}(I, \bar{\Theta}_2^2),$
	Relations	$\bar{\Gamma} = \begin{bmatrix} I & 0 \\ 0 & 0 \\ 0 & \bar{\Gamma}_{3,2} \end{bmatrix}$
Stationary subproblem	Formula	$\langle \tilde{z}_{t,3}^{\pi}, \tilde{y}_{t,2}^\pi\rangle\langle \tilde{y}_{t,2}^\pi, \tilde{y}_{t,2}^\pi\rangle^{-1}\langle \tilde{y}_{t,2}^\pi, \tilde{z}_{t,3}^{\pi}\rangle\bar{\Gamma}_{3,2} = \langle \tilde{z}_{t,3}^{\pi}, \tilde{z}_{t,3}^{\pi}\rangle\bar{\Gamma}_{3,2}\bar{\Theta}_2^2$
	Restrictions	$\bar{\Gamma}'_{3,2}S_{p,22} = I$
	Relations	converges to $\tilde{b}_{2,3} = \tilde{O}_2\Gamma'_{3,2}$.

Table 1: Singular value decompositions used in the article.

$$\bar{\Psi} = \begin{bmatrix} T^{-1}\langle \tilde{z}_{t,1}^\pi, \tilde{z}_{t,1}^\pi\rangle & T^{-1}\langle \tilde{z}_{t,1}^\pi, \tilde{z}_{t,2}^\pi\rangle & 0 \\ T^{-1}\langle \tilde{z}_{t,2}^\pi, \tilde{z}_{t,1}^\pi\rangle & T^{-1}\langle \tilde{z}_{t,2}^\pi, \tilde{z}_{t,2}^\pi\rangle & 0 \\ 0 & 0 & \langle \tilde{z}_{t,3}^\pi, \tilde{z}_{t,3}^\pi\rangle \end{bmatrix}$$

$$J = \bar{Q} - \bar{\Phi}.$$

$$Z_{ij} = T^{-1}\langle \tilde{z}_{t,i}^\pi, \tilde{z}_{t,j}^\pi\rangle, j = 1, 2.$$

$$\delta_{zz} = \bar{M} - \bar{\Psi}$$

$$\delta_{yz} = \begin{bmatrix} T^{-1}(\langle \tilde{y}_{t,1}^\pi, \tilde{z}_{t,1}^\pi\rangle - \langle \tilde{z}_{t,1}^\pi, \tilde{z}_{t,1}^\pi\rangle) & T^{-1}(\langle \tilde{y}_{t,1}^\pi, \tilde{z}_{t,2}^\pi\rangle - \langle \tilde{z}_{t,1}^\pi, \tilde{z}_{t,2}^\pi\rangle) & T^{-1/2}\langle \tilde{y}_{t,1}^\pi, \tilde{z}_{t,3}^\pi\rangle \\ T^{-1/2}\langle \tilde{y}_{t,2}^\pi, \tilde{z}_{t,1}^\pi\rangle & T^{-1/2}\langle \tilde{y}_{t,2}^\pi, \tilde{z}_{t,2}^\pi\rangle & 0 \end{bmatrix}$$

$$\delta_{yy} = \begin{bmatrix} T^{-1}(\langle \tilde{y}_{t,1}^\pi, \tilde{y}_{t,1}^\pi\rangle - \langle \tilde{z}_{t,1}^\pi, \tilde{z}_{t,1}^\pi\rangle) & T^{-1/2}\langle \tilde{y}_{t,1}^\pi, \tilde{y}_{t,2}^\pi\rangle \\ T^{-1/2}\langle \tilde{y}_{t,2}^\pi, \tilde{y}_{t,1}^\pi\rangle & 0 \end{bmatrix}$$

$$\bar{S} = (\langle \tilde{z}_{t,3}^\pi, \tilde{z}_{t,3}^\pi\rangle - \langle \tilde{z}_{t,3}^\pi, \tilde{y}_{t,2}^\pi\rangle\langle \tilde{y}_{t,2}^\pi, \tilde{y}_{t,2}^\pi\rangle^{-1}\langle \tilde{y}_{t,2}^\pi, \tilde{z}_{t,3}^\pi\rangle)^{-1}.$$

$$\bar{\Gamma}^\dagger = (\bar{\Gamma}'\bar{\Psi}\bar{\Gamma})^{-1}\bar{\Gamma}$$

$$\bar{G}^\dagger = (\bar{G}'\bar{M}\bar{G})^{-1}\bar{G}$$

$$\delta G = \bar{G} - \bar{\Gamma}$$

$$\tilde{O}_2^\dagger = (\tilde{O}_2'(\mathbb{E}\tilde{y}_{t,2}^\Pi\tilde{y}_{t,2}')^{-1}O_2)^{-1}\tilde{O}_2'(\mathbb{E}\tilde{y}_{t,2}^\Pi\tilde{y}_{t,2}')^{-1}.$$

B.3 Singular value decompositions